PROCEEDINGS
OF THE
NATIONAL ACADEMY OF SCIENCES

Volume 12 FEBRUARY 15, 1926 Number 2

THE RED SHIFT OF SOLAR LINES AND RELATIVITY

By Charles E. St. John

Mount Wilson Observatory of the Carnegie Institution of Washington

Read before the Academy, April 24, 1924

The differences in wave-length between the center of the sun's disk and the vacuum arc for 331 iron lines have been used as the basis of a discussion of the displacements of the solar lines in relation to the theory of relativity.

In favor of relativity is the general displacement to the red, and against it, certain discordances between the observed and theoretical values. The third and seventh columns of table 1 illustrate what has long been recognized, namely, that both strong and weak lines give displacements at the center of the disk which differ systematically from the calculated values, while lines of medium intensity show displacements in substantial agreement with the theoretical requirement. The observed displacements of strong lines are about 50 per cent greater, and of weak lines about 30 per cent less than the predicted results. This discrepancy has long stood in the way of an interpretation based upon general relativity. It is the purpose of the present discussion to suggest a harmonizing interpretation.

TABLE 1

Comparison of Observed Displacements (Sun *minus* Vacuum) with Those Calculated from the Theory of Relativity

	No. of Group Lines	Solar Intensity	Mean Wave-length	Calculated	Observed	Obs.-Cal.	Radial Velocity
a	17	12	3826 A	0.008 A	0.012 A	+0.004 A	0.3 km./sec. down
b	24	14	3821	0.008	0.0112	+0.0032	0.25 km./sec. down
b	10	10.4	4308	0.0091	0.0113	+0.0022	0.16 km./sec. down
a	10	6.2	5443	0.0115	0.0112	−0.0003	.....
b	131	4.8	4758	0.0100	0.0084	−0.0016	0.1 km./sec. up
d	106	4.5	4763	0.0100	0.0069	−0.0031	0.2 km./sec. up
a	33	3.3	4957	0.0105	0.0074	−0.0031	0.2 km./sec. up

The key to the solution offered is that the deviations from theory shown by the strong and weak lines are correlated primarily, not with line-intensity, but with the level in the sun's atmosphere at which the lines

are produced. This is clearly evident from the behavior of lines of enhanced and normal titanium as shown in table 2. Lines in the flash spectra at heights of 6000, 1300 and 435 km. show absolute displacements of +0.015, +0.0112 and 0.0054 A, respectively, and deviations of +0.007, +0.002 and −0.0034 A from the displacements calculated according to the relativity theory, a definite correlation with level. That the deviations are not fundamentally associated with line-intensity is shown by comparing the last two groups, consisting of enhanced and normal lines of approximately equal intensities, which show absolute displacements of +0.0112 and +0.0054 A, and deviations of +0.002 and −0.0054, respectively, from the calculated values.

TABLE 2

RED-DISPLACEMENT AND LEVEL

	NO.	MEAN INT.	SUN-VAC.	OBS.-CAL.	HEIGHT
Enh. Ti	2	11	+0.015 A	+0.007 A	6000 km.
Enh. Ti	8	4.6	+0.0112	+0.002	1300
Normal Ti	5	4.2	+0.0054	−0.0034	435

The systematically larger displacement of high-level lines toward the red which appears clearly in table 1, has also been observed on high-dispersion spectrograms of Sirius, Procyon and Arcturus, and, in general, is the more pronounced the higher the temperature of the star.

In attempting to interpret the observed displacements, it is to be remarked that the low pressure now known to exist in the reversing layer of the sun, and confirmed by independent observations on stars, removes at one stroke the necessity of considering pressure shift and ray-curving due to anomalous refraction; further, that the Zeeman effect of the sun's general magnetic field produces at most only a slight symmetrical widening of the lines, and that the Stark effect has not been found in the sun, although looked for under conditions apparently favorable to its appearance.

There remain to be considered apparently only the Doppler effect and general relativity, neither of which alone is capable of producing red-displacements of the solar lines in agreement with the observations. In combination, however, these two offer a simple interpretation of the displacements at the center of the disk. Relativity provides for the general displacement, while the Doppler effect, on the basis of apparently good evidence, can be invoked in explanation of the systematic deviations of high and low level lines. The last column of table 1 shows the magnitudes and the directions of the radial movements in the solar vapors that would be required to account for these deviations from Einstein's theory.

The existence of radial movements in the solar atmosphere is inherently probable, and has, moreover, the support of observational evidence. Thus a study of the K lines of calcium in 1910[1] showed that at high levels the

vapor producing the K_3 line is descending with a velocity of 1.14 km./sec. when the whole displacement between the center of the sun and the arc was attributed to motion, while the lower level of the K_2 line is an upward motion of 1.97 km./sec. These values refer to the center of the disk, but it is noteworthy that, as the limb is approached, the displacements decrease in close agreement with the cosine law. There can be little doubt, therefore, that the interpretation based on the hypothesis of vertically descending and ascending currents is the correct one.

The much smaller velocities required to explain the deviations in table 1 would appear to be the result of a spectrographic integration of the light from extended areas of the solar surface. At high levels the absorption produced by the cooler downward-drifting vapors is of dominating influence, and the integrated effect is an asymmetry on the red side, increasing with increasing elevation. At low levels the rapidly rising currents over the granulations will be more effective in producing the lines than the slower downward currents over larger and cooler areas, with a resulting displacement toward the violet. The decrease of upward velocity with elevation brings about a balanced state for lines of medium intensity, produced at intermediate levels.

TABLE 3

LIMB-CENTER DISPLACEMENTS (ADAMS)

REGION	INTENSITY 1	2–3	4–6	STRONGEST LINES
4300	+0.0038	+0.0052	+0.0062 A	
4800	+ .0049	+ .0062	+ .0077	Very small
5300	+ .0062	+ .0072	+ .0087	Practically zero
5800	+ .0075	+ .0085	+ .0093	
6300	+ .0087	+ .0092	+ .0105	

As the limb is approached the systematic difference between high and low level disappears, just as in the case of the K line of calcium. When the limb-center displacements shown in table 3 are added to the displacements at the center, the red displacements of essentially all lines exceed the relativity of displacements by small amounts. This excess, the real limb effect from the point of view of relativity, may be interpreted as the result of molecular scattering in accordance with the Rayleigh-Schuster formulæ. Professor Julius has called attention to the fact that, in general, the refractive power is greater on the red than on the violet side of an absorption line by twice the refractive power of the solar atmosphere. Since the coefficient of scattering increases as the square of the refractive power, the differential scattering tends to widen the lines on the red edge. The short paths through layers of low density at the center of the disk would account for the absence of differential widening at the center, while the greatly lengthened paths at the limb would furnish conditions favorable to its appearance.

The conclusion is that three major causes produce the differences between solar and terrestrial wave-lengths, and that it is possible to disentangle their effects. These appear to be the slowing up of the atomic clock in the sun to an amount predicted by the theory of generalized relativity, radial velocities of moderate cosmic magnitude and in probable directions, and differential scattering, which is most effective in the longer paths through the solar atmosphere traversed by light coming from the limb of the sun. The first obtains for all lines in all parts of the sun, the second, active downward at very high and upward at very low levels, is most evident at the center of the disk and vanishes at the limb; while the third manifests itself in the so-called limb effect.

[1] *Mt. Wilson Contr.*, No. **48**; *Astrophysical Journal, Chicago, Ill.*, **32**, 36–82, 1910.

SOME RELATIONS OF ENVIRONMENT TO THE EPIDEMIOLOGY AND CONTROL OF APPLE SCAB[1]

BY G. W. KEITT

COLLEGE OF AGRICULTURE, UNIVERSITY OF WISCONSIN

Read before the Academy November 11, 1925

Apple scab, which is one of the most destructive diseases of fruits, varies greatly from year to year both in severity of occurrence and difficulty of control. What is the explanation of these variations, and how may more adequate control measures be developed? Light on these questions has been sought through a study of the relations of environment to the epidemiology and control of the disease. This work has been directed along two lines: (1) field studies of the relation of the natural environment to the development and control of the disease, and (2) laboratory and greenhouse studies of the development and prevention of the disease under conditions in which certain factors of the environment were controlled.

Apple scab is caused by *Venturia inaequalis* (Cke.) Winter, an ascomycetous fungus which parasitizes the fruit, leaves and, more rarely, the twigs of its host. Under Wisconsin conditions, this fungus overwinters in infected leaves which fall to the ground. In spring, under favorable conditions, ascospores are forcibly ejected from perithecia developed in these leaves, and are carried by air currents to susceptible host parts, where they induce primary infection. Successive generations of conidia borne on infected parts spread the disease throughout the season. These cardinal points of life history have long been known, and have contributed to the basis for control programs, which at present depend in a minor degree upon disposal of infected leaves and to a major extent upon pro-

tection of susceptible parts by fungicidal applications. Why, however, do we find moderate scab development and successful control in one year or in one locality, and an epidemic outbreak and inadequate control in another?

I. FIELD STUDIES.—Field studies were conducted with a four-fold purpose: (1) to procure detailed records pertaining to the development and control of the disease in relation to certain factors of the natural environment, (2) to solve certain urgent practical problems without the delay incident to the development of apparatus and technique for work under controlled conditions, (3) to define problems for laboratory and greenhouse study under controlled conditions, and (4) to check against the results of experiments in which certain factors of the environment were controlled.

To facilitate these studies a meteorological station equipped with apparatus of the types used by the U. S. Weather Bureau was established in the experimental orchards. From this station are obtained hourly records of temperature, rainfall, wind velocity, relative humidity and periods of bright sunshine. These data are correlated with a series of records which have been developed for tracing certain significant aspects of the seasonal development of the fungus, the apple plant, the disease and the effectiveness of control measures.

These field records, which cannot here be discussed in detail, have served their purposes very satisfactorily. They have shown, in general, agreement with many previous observations by other investigators, that the moisture and temperature factors of the natural environment play a leading rôle in determining the severity of the disease and the difficulty of its control. They have defined critical points in epidemiology and control, and have led to a material increase in the efficiency of the fungicidal program through its more effective orientation to critical periods.

II. EXPERIMENTS UNDER CONTROLLED CONDITIONS. (1) *Apparatus and Method.*—Through the kindness of Doctors J. G. Dickson[2] and James Johnson,[3] their apparatus for the control of air temperature and humidity was made available for our use. This equipment was very useful for studies during the period following the initial stages of infection. For studies during the early stages of spore germination and infection, however, it was necessary to devise a chamber (to be described in a later paper) in which a saturated atmosphere can be maintained at a wide range of constant temperatures. Potted apple plants of suitable size were placed in this apparatus, and inoculated by abundant natural discharge of ascospores of *V. inaequalis* from moist overwintered infected leaves which were suspended on trays of wire netting in the upper part of the chamber. Condensation water on the susceptible parts of the host furnished an adequate supply of moisture for spore germination and infection. This method

provides an inoculum which is remarkably free from extraneous material, and which, in an unusual degree, parallels natural conditions.

(2) *Relations of Temperature.*—In studying the relations of temperature to infection, it has seemed desirable to differentiate between the initial stages of infection, up to the time when the parasite becomes independent of an external water supply, and the remaining stages of disease development.

(*A*) *During Initial Stages in Moist Chamber.*—The experimental plants were inoculated and held in the moist chamber at various constant temperatures (variable less than 1°C.) ranging from 6 to 30°C. Infection resulted (table 1, Series I) at temperatures ranging from 6° to 26°C. It is probable that further studies will somewhat extend this range, particularly toward the lower limit. A further criterion of the relation of temperature to the initial stages of infection consists in the time required for the fungus to establish itself in the host tissues sufficiently to become independent of an external moisture supply. The results of experiments

TABLE 1

REPRESENTATIVE RESULTS FROM EXPERIMENTS ON THE RELATION OF TEMPERATURE AND MOISTURE TO APPLE LEAF INFECTION BY ASCOSPORES OF THE APPLE SCAB FUNGUS

SERIES AND DATES OF INOC. 1925	INOC. CHAMBER TEMP. °C.	CHAMBER PERIOD IN HRS.	INCUBATION CONDITIONS[a]	INCUBATION PERIOD DAYS	RESULTS—AVERAGED PER TWIG NO. LEAVES INFECTED[b]	RESULTS—AVERAGED PER TWIG MAX. NO. LESIONS ON ONE LEAF
Series I						
Apr. 2	6	24	20–25°C.	11	2.5	11
Apr. 13	9	do.	do.	10	3.1	19
Apr. 10	15	do.	do.	9	3.2	27
Apr. 30	20	do.	do.	11	3.3	40
Apr. 20	24	do.	do.	14	2.8	22
Mar. 27	26	do.	do.	13	1	2.4
May 28	28	do.	do.	..	0	0
Series II						
Apr. 2	6	13	do.	..	0	0
Do.	6	18	do.	11	2.2	5
Do.	6	24	do.	11	2.5	11
Do.	6	36	do.	11	2.9	21
Do.	6	44	do.	11	3.1	26
Series III						
Apr. 8	20	21	do.	9	2.9	15
Do.	20	do.	20°C.–90% R. Hum.	9	3	18
Do.	20	do.	20°C.–80% R. Hum.	9	2.9	16
Do.	20	do.	20°C.–50% R. Hum.	9	3.1	15

[a] In Series I and II, in which the variations in treatments were made in the inoculation chamber, the plants were incubated at 20–25°C. in a greenhouse in which no effort was made to control humidity. In Series III, in which the treatments were varied after inoculation, the plants were incubated in specially designed chambers (p. 69) at 20°C. and relative humidities of 50%, 80% and 90%, respectively.

[b] The older leaves become resistant to scab infection.

on this aspect of infection appear in table 2. Infection occurred freely at 6°C., the lowest temperature tried, although the process was much retarded. The periods necessary for infection were progressively shortened in trials at 9°, 15° and 20°C. The results at 24° duplicated those at 20°, while at 26° the period was considerably lengthened. These data indicate an optimal rate of progress of the initial stages of infection at temperatures near 20°C.

TABLE 2

RESULTS OF EXPERIMENTS ON THE MINIMAL PERIODS OF CONTINUOUS WETTING NECESSARY AT VARIOUS TEMPERATURES FOR APPLE LEAF INFECTION BY ASCOSPORES OF THE APPLE SCAB FUNGUS

TEMPERATURE °C.	PERIODS OF CONTINUOUS WETTING WHICH GAVE: NEGATIVE RESULTS HRS.	POSITIVE RESULTS HRS.
6	13	18
9	11	13
15	..	8.5
20	4	6
24	4	6
26	8	10

(*B*) *During Incubation.*—From limited data it appears that the effect of temperature during the incubation period parallels its influence upon the initial stages of leaf infection. At 8°C. the period of incubation was prolonged to 17 days, as compared with periods commonly ranging from 8 to 12 days at temperatures of 20–25°C. When the temperature during incubation was raised to 26°C., the disease developed sparsely after 13 days in one experiment and failed to develop in two. It is evident, however, that in nature the organism frequently tolerates periods of summer heat in which temperatures materially exceed 26°C. The effect of intermittent exposures to higher temperatures is, therefore, of interest. Four intermittent exposures of 8 hours each at 31°C. during the period of incubation did not preclude the development of the disease, although they appeared to inhibit it somewhat. A single exposure of 24 hours to 31–32°C. did not preclude scab development. Similar exposures of 48 hours or more, however, prevented macroscopic development of lesions.

(*C*) *To Parasite and Host.*—Figure 1 shows the results of a representative experiment on germination of ascospores at constant temperatures. It will be observed that the relations of temperature to ascospore germination closely parallel thermal relations to infection. In each case the maximal rate of development is near 20°C. While the thermal range for ascospore germination is slightly greater than the demonstrated range for infection, it should be recalled that further work may show a slight extension of the temperature range for infection, particularly at the lower limit. As yet little detailed data are available on the responses of the host plant to temperature. It is evident, however, that its temperature range for active

growth is distinctly higher than that for growth of the fungus or development of the disease. These data, while still fragmentary, accord well with the results of field studies in showing that *V. inaequalis* is an active parasite at low and intermediate temperatures, but is readily checked by moderate heat.

(3). *Relations of Moisture.* (*A*) *The Minimal Period of Continuous Wetting Necessary for Infection.*—In studies of epidemiology and control it is of much importance to know the minimal periods of continuous wetting necessary for infection at various temperatures. Data bearing

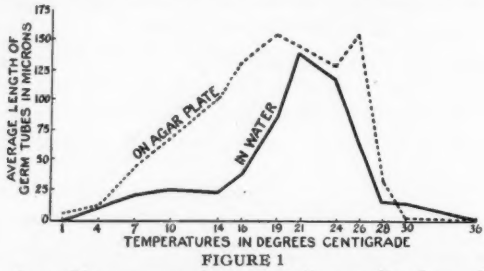


FIGURE 1

A graphic summary of representative results of experiments on the relation of temperature to the germination of ascospores of the apple scab fungus.

on this subject appear in table 1, Series II, and table 2. For reasons which cannot be discussed here because of limitations of space, it is apparent that most natural infections require somewhat longer periods of wetting than the minimal periods indicated in the tables.

(*B*) *Discontinuous Wetting in Relation to Infection.*—Aderhold early showed that ascospores of *V. inaequalis* germinate well if exposed to alternate wetting and drying, and opined that discontinuous wetting actually favors infection by stimulating the formation of appressoria, which play an important rôle in the mechanism of penetration of the fungus. The present writer's experiments have shown that during the early stages of infection the fungus can readily tolerate brief periods of drying and of exposure to direct sunlight. Field studies, however, indicate that fractional or intermittent infection periods play a less important rôle in nature than these results might suggest.

(*C*) *Humidity in Relation to Infection.*—In considering the relations of humidity to infection, at least three aspects should be taken into account: (*a*) its effect upon the host plant prior to infection, particularly in relation to the development of the cuticle, (*b*) its influence upon spore germination and the initial stages of infection, and (*c*) its effect during the period of incubation. The data now available apply only to the last of these phases. The results of a representative test appear in table 1, Series III. There is no

evidence that these experimental variations in humidity exerted any significant influence upon the development of the disease. These data are too fragmentary to be more than suggestive. They accord with field observations, however, in suggesting that ordinary fluctuations in humidity will not significantly affect the progress of the disease during the period of incubation.

(4). *The Effectiveness of Fungicides in Relation to Environment.*—The literature of spraying and dusting contains many conflicting reports concerning the effectiveness of fungicides. The reasons for these discrepancies are too complex for analysis here. One potentially important consideration, however, is the influence of environmental factors, particularly temperature and moisture, upon the effectiveness of fungicidal treatments. While numerous field observations have been made concerning these relationships, little work has been done under conditions in which the fungicidal applications could be related at will to the experimental production of the disease under controlled conditions. Such studies, which were made possible by the apparatus and method of inoculation described above, are in progress. The most noteworthy results obtained thus far are (1) that commercial lime-sulphur and sulphur dust, in the concentrations ordinarily used in practice, prevented apple leaf infections by ascospores of *V. inaequalis* at 6°C., the lowest temperature tried, and (2) that, *under the conditions of these experiments*, the sulphur fungicides named were consistently more effective than the standard copper spray, Bordeaux mixture (4–4–50).

III. DISCUSSION.—The results herein reported may be of interest primarily as they constitute a small accretion to a body of knowledge which is gradually broadening our conception of the nature and control of parasitic diseases. The time has long passed when parasitic diseases could be regarded as essentially fixed or stable. Developments in genetics, physiology, pathology and other fields have established the fact that hosts and parasites alike are plastic, each potentially variable through the play of heredity on the one hand and environment on the other. Consequently, in plant pathology at least, we are forced increasingly to take into account the significance of variability, not only as affecting host and parasite, but in relation to epidemiology and control. Insofar as the present writer's studies have progressed, they indicate that the marked variations which commonly occur in the development and control of apple scab are to be attributed largely to fluctuations in moisture and temperature acting chiefly in the following relationships: (1) to the development of the causal organism, particularly by influencing the timeliness and abundance of spore production and the conditions for infection at critical stages of host development; (2) to the development of the host, particularly by influencing the time required to pass through its stages of maximal susceptibility; and (3) to the efficiency of fungicides.

A more detailed account of this work is being prepared for publication.

[1] For some years investigations of the relations of environment to plant disease have been in progress at the University of Wisconsin (see Jones, L. R., "The Relation of Environment to Disease in Plants," *Amer. J. Bot.*, **11**, 601–609 (1924)). The present paper treats of a disease which affects only certain aerial parts of the host plant, and deals primarily with the relations of the aerial environment to the development and prevention of this disease.

Acknowledgments are made to Dr. Leon K. Jones as collaborator in much of the work herein discussed.

[2] An account of Dr. Dickson's apparatus will appear in a forth-coming Research Bulletin of the Wisconsin Agricultural Experiment Station.

[3] Johnson, J., "The Relation of Air Temperature to Certain Plant Diseases," *Phytopathology*, **11**, 446–458 (1921).

INFRA-RED ABSORPTION OF THE N—H BOND

By E. O. Salant*

Department of Physics, Johns Hopkins University

Communicated January 9, 1926

Introduction.—From his examination of the infra-red spectra of nitrogenous compounds, Coblentz showed that an absorption band at $\lambda = 6.1$ to 6.2μ and its first harmonic at $\lambda = 2.96\mu$ are characteristic of the NH_2 radical.[1]

From certain considerations of structure and heat capacities of organic substances, to be treated by the writer in another paper, it seemed likely that the absorption band characteristic of the NH_2 radical is caused by the N—H bond. The measurements presented in this paper were made to test that conclusion.

In his very extensive investigation of the substituted anilines, Bell,[2] using a rock-salt prism, found a band from 2.7 to 2.9μ present in aniline and the mono-alkyl anilines, but practically absent in the dialkyl anilines. Since, of course, aniline and the mono-alkyl anilines have N—H bonds, whereas the dialkyl anilines have not, this may be interpreted to mean that the 2.7 to 2.9μ band is due to the N—H bond. Unfortunately, however, the absence of the band in the substances without N—H is not quite certain, two of the four dialkyl anilines examined having slight absorption bands around 2.9μ.

Accordingly, a series of symmetrical dialkyl and trialkyl amines was selected for examination in the region of 3μ. For, if a band in that region is due to the N—H bond, the transmission curves of the dialkyl amines should all show minima there, whereas, unless there is some other configura-

tion in their molecules causing absorption at 3μ, the curves of the trialkyl amines should not.

The substances examined are: di-n-propyl amine, tri-n-propyl amine, di-n-butyl amine, tri-n-butyl amine, di-iso-amyl amine, tri-iso amyl amine, di-phenyl amine, tri-phenyl amine, di-benzyl amine, tri-benzyl amine, saturated benzene solutions of azobenzene and of hydrazobenzene, and also toluidine and dimethyl toluidine. Except for the two solutions and for the phenyl amines and tri-benzyl amine, which are solids, the compounds were examined in the pure liquid state. All were anhydrous, Eastman Kodak Co. "highest purity."

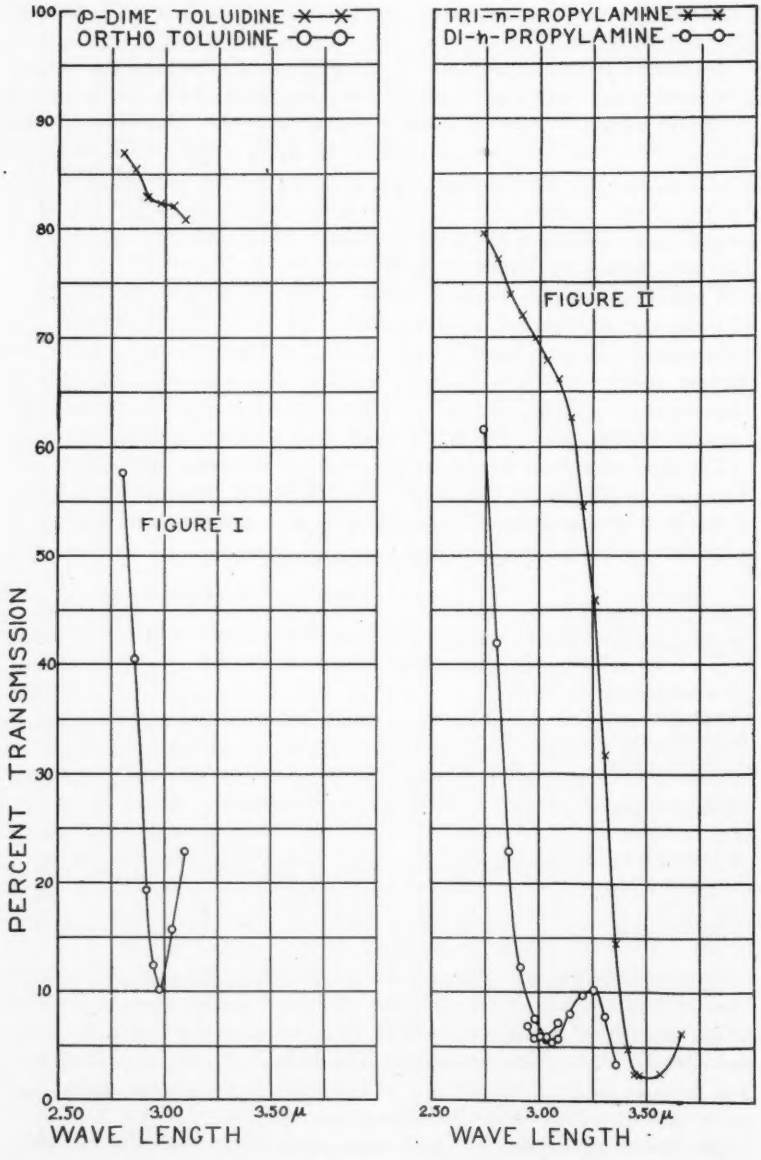
A gas-filled 500-watt tungsten lamp, served by a steady current from 120-volt storage batteries was the source. The Wadsworth spectrometer, with mirrors 10 cm. diameter and 52 cm. focal length, and a 60° clear fluorite prism with faces 33 mm. in diameter were used. A Coblentz bismuth-silver thermopile and Thomson galvanometer measured the transmitted radiation. These have all been already adequately described.[3]

The absorption cells for the liquids, of the same design as the rock-salt cells used by Coblentz,[4] consisted of fluorite plates separated by a frame of tin-foil 0.05 mm. thick. The solids were merely melted and allowed to crystallize between fluorite plates.

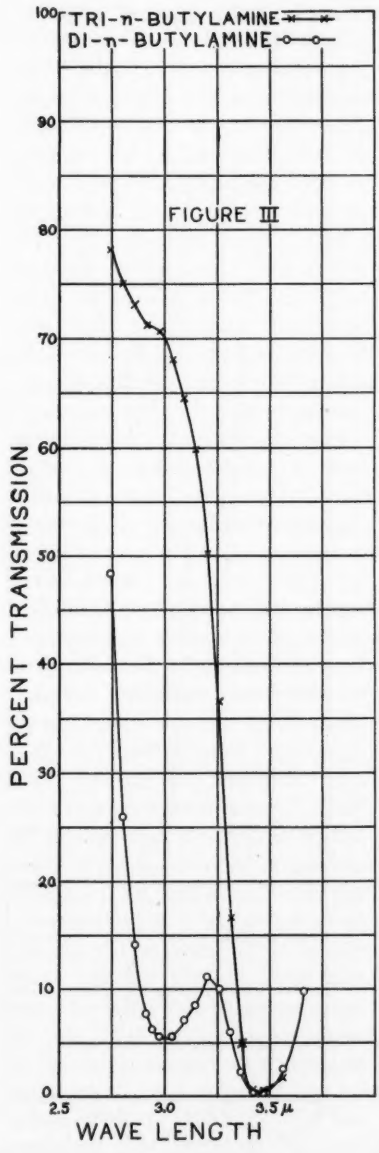
SUBSTANCE	WAVE-LENGTH IN μ OF TRANSMISSION MINIMA		
Tri-n-propylamine			3.48–3.52
Di-n-propylamine	3.04		
	3.00–3.08		
Tri-n-butylamine			3.45
Di-n-butylamine	3.01		3.47
Tri-iso-amylamine			3.47
Di-iso-amylamine	3.07		
	3.02–3.10		3.46
Tri-phenylamine		3.26	
Di-phenylamine	2.89	3.23	
Tri-benzylamine		3.315	3.475
Di-benzylamine	3.00	3.315	3.47
Azobenzene			
Hydrazobenzene	2.99		

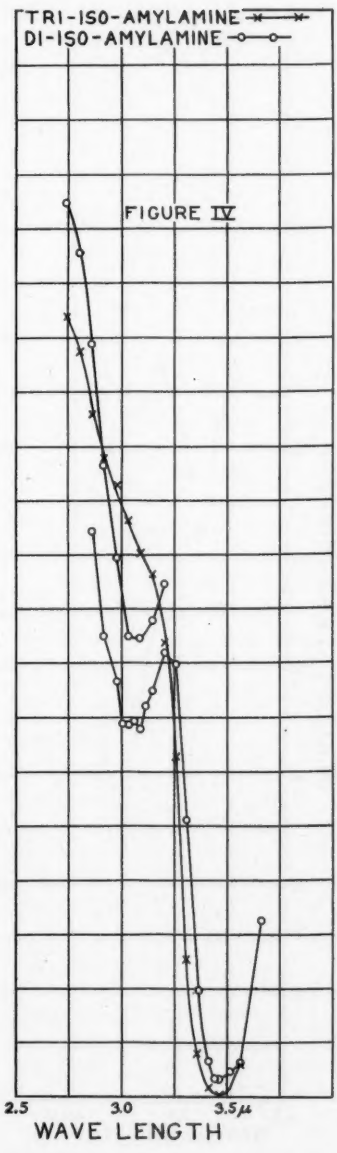
Discussion of Results.—Figure 1 shows a deep absorption band for toluidine at 2.98μ, but none for dimethyl toluidine—merely another example of the absorption of the NH_2 radical. The band of the C—H bond[5] appears from 3.45 to 3.52μ in the aliphatic amines, figures 2, 3, 4 and 6. The benzene band[1] is evident at 3.23 and 3.26μ in the phenyl amines, figure 5, and at 3.315μ in the benzylamines, figure 6.

The following substances show absorption from 2.9 to 3.1μ: di-n-propyl amine, di-n-butyl amine, di-iso-amyl amine, di-phenyl amine, di-benzyl amine and hydrazobenzene. The following substances show

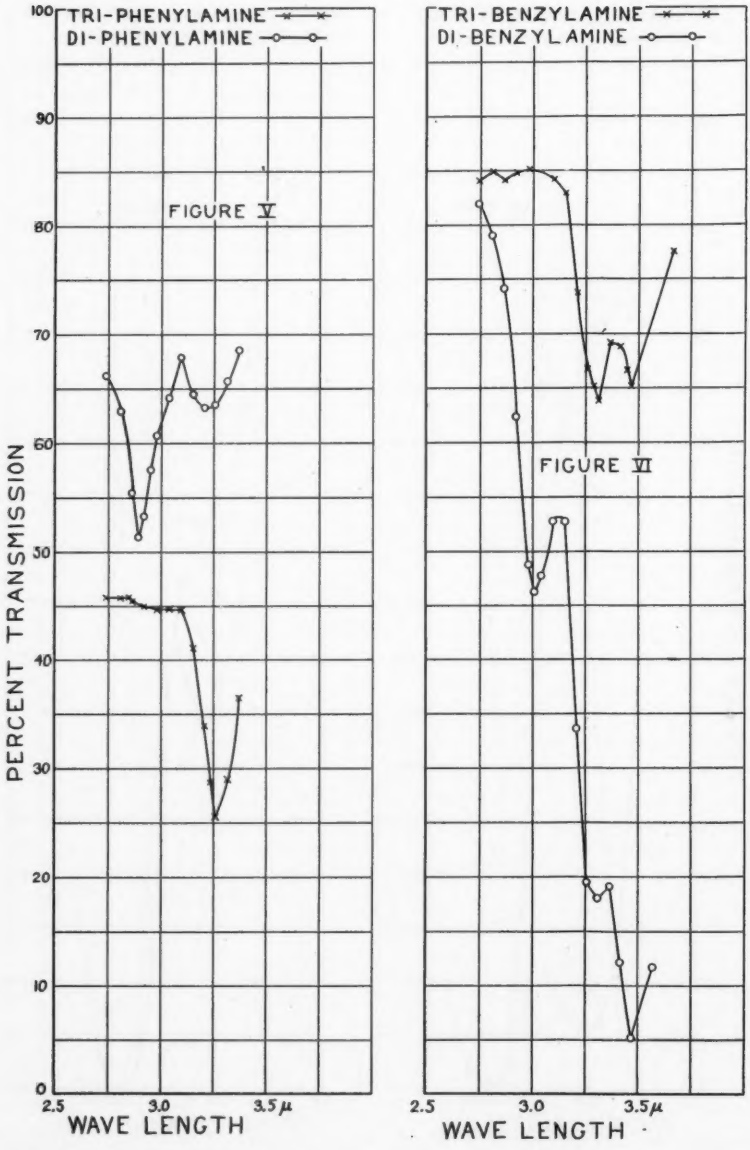


no absorption from 2.9 to 3.1μ: tri-n-propyl amine, tri-n-butyl amine, tri-iso-amyl amine, tri-phenyl amine, tri-benzyl amine and azobenzene.





It is obvious, therefore, that an absorption band in the region of 3μ is characteristic of the N—H bond. The variation of the band within that



region for each particular substance is shown by the tabulated transmission minima.

It is to be noted that, whereas the band of the N—H bond in the other dialkyl amines is between 3.0 and 3.1μ, it is at 2.89μ in diphenyl amine. As this is the only solid examined having the N—H bond, it might be thought that the shift of the band to the shorter wave-length is due to a stronger binding resulting from the solid state, but as the bands in the solid ammonium salts are at 3.2μ,[6] that explanation is not satisfactory. The small influence of physical state on the position of the bands is brought out quite clearly by the absorption of *dissolved* hydrazobenzene at 2.99μ, figure 7. The shift of the band in di-phenyl amine is more likely related to the binding of the nitrogen to an unsaturated carbon atom.

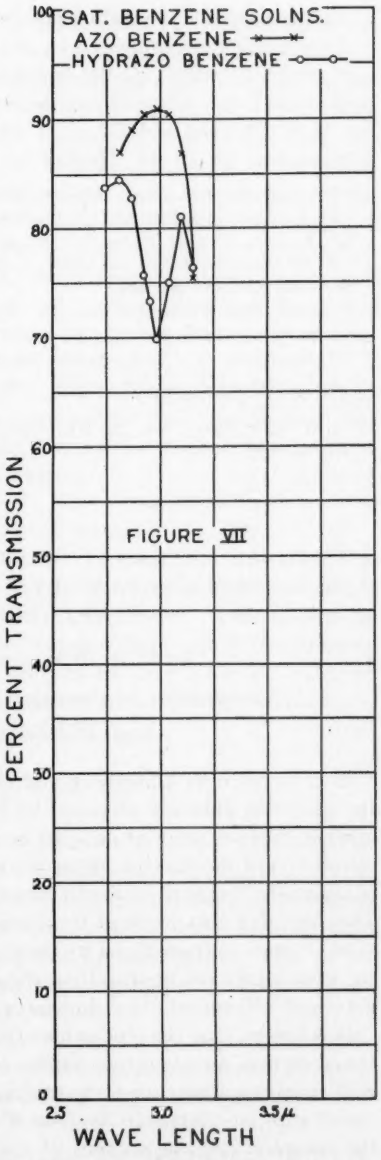


FIGURE VII

Kratzer[7] first showed that certain frequencies in gases can be taken as approximate harmonics of a fundamental frequency, according to the equation $\nu = n\nu_0(1 - nx)$, where ν_0 is the fundamental frequency, $n = 2$ for the first harmonic, etc., and x is a correction factor. Ellis[8] has shown that this applies to the bands of the C—H bond in organic liquids. Now, the ammonium ion in ammonium salts has a characteristic absorption from 6.69 to 7.15μ[6] (wave number 1495 to 1400) and ammonia itself at 6.1μ (wave number 1640). If, then, these are taken as the fundamentals, of which the frequencies in the region of $\lambda = 3\mu$ (wave number 3333) are the approximate first harmonics, the results of this paper indicate that the N—H bond has a characteristic fundamental vibration frequency varying between wave numbers 1400 and

1700, the variation depending mainly on the other atoms in the molecule.

I am particularly indebted to Dr. W. W. Coblentz for his suggestions and personal interest in the performance of these measurements, which were made in his laboratory in the Bureau of Standards. I wish to thank Dr. G K. Burgess and Dr. C. A. Skinner for permission to work at the Bureau, and Mr. C. W. Hughes for aid in the laboratory.

* Johnston Scholar, Johns Hopkins University.

[1] W. W. Coblentz, *Carnegie Inst. Washington Pub.*, No. **35**, Part II, p. 117.

[2] F. K. Bell, *J. Amer. Chem. Soc.*, **47**, 2192 (1925).

[3] W. W. Coblentz, *J. Opt. Soc. Amer.*, **7**, No. 6 (1923); Bureau of Standards, *Scientific Papers* **188, 204, 282** and **401.**

[4] *Carnegie Inst. Washington Pub.*, No. **35**, p. 24.

[5] J. W. Ellis, *Physic. Rev.*, [2] **23**, 48 (1924).

[6] O. Reinkober, *Zs. Physik.*, **5**, 192 (1921).

[7] A. Kratzer, *Ibid.*, **3**, 289 (1920). See also Sommerfeld, *Atomic Structure and Spectral Lines*, Ch. VII, § 2.

[8] J. W. Ellis, *Physic. Rev.*, **22**, 200 (1923).

A POSSIBLE EXPLANATION OF THE RELATIVITY DOUBLETS AND ANOMALOUS ZEEMAN EFFECT BY MEANS OF A MAGNETIC ELECTRON

By F. R. Bichowsky and H. C. Urey

Department of Chemistry, Johns Hopkins University

Communicated December 22, 1925

In a recent note Uhlenbeck and Goudsmit[1] suggest that the fourth of the quantum numbers proposed by Pauli[2] to describe the quantum state of each electron in an atom is to be associated with the angular moment of rotation of the electron about its own axis. The electron thus assumed is essentially identical except in dimensions with the magneton of Parsons.[3] This idea had also occurred to us quite independently and for largely the same reasons as those given by these authors. We have, however, carried the idea somewhat further than the authors have reported in their brief note and will report these further considerations here.

It is known that the separation of the members of the multiplet spectral terms such as are concerned in the *D* lines of sodium must be connected with magnetic forces since the multiplicity of the term is invariably associated with an anomalous Zeeman effect and since at high magnetic fields the complex Zeeman pattern of the multiplet term simplifies into the normal Zeeman triplet (the Paschen Back effect). It was therefore proposed by Heisenberg[4] that the multiplet separation is due to the difference

in the mutual magnetic energy of the atom when the valence electrons were moving in the same sense as the resultant motion of the rest of the atom and when they were moving in whole or in part in the opposite sense to the supposed resultant motion of the rest of the atom. This model was calculated by Heisenberg on the simplifying assumption that the magnetic force on each valence electron could be calculated from the vector sum of the external field and the resultant field due to the relative motion of the kernel. The result was shown to be consistent with the observed Zeeman and Paschen-Back effect of the normal doublets and triplets but as was pointed out by Breit[5] gave an energy order for the p_1 and p_2 and analogous level opposite to that observed. We noticed that this error in sign would be corrected if the electron were rotating on itself and moving about a positive kernel of no magnetic moment.

It was shown by Millikan and Bowen[6] and also by Landé[7] that these optical doublets and triplets obey the same law for their separation as the so-called relativity X-ray doublets, that is to say that the energy separation of all of these levels is given by the formula

$$\Delta W = \frac{8\pi^4 e^8 m}{c^2 h^4 n^3 k_1 k_2} Z_i^2 Z_a^2 \Delta m,$$

where $k_1 - k_2 = 1$ and where Z_i and Z_a are the effective atomic numbers for the inner and outer parts of the orbit, respectively. This is identical with the formula originally proposed by Sommerfeld for the X-ray doublets on relativity considerations except that in his case $Z_i = Z_a$. Landé[8] attempted the explanation of these optical and X-ray doublets on the basis of a magnetic kernel and secured the formula

$$\Delta W = \frac{8\pi^4 e^8 m}{c^2 h^4 n^3 k^2} Z_i Z_a^2 \Delta m,$$

where Δm is the difference of the two quantum numbers specifying the orientation of the orbit to the axis of the kernel and is equal to 1. This formula has the same constant as Sommerfeld's formula but is proportional to Z_i instead of Z_i^2, or Z_i^3 instead of Z_i^4 for the X-ray levels, as is the case in Sommerfeld's formula and as is required by experiment. This model can therefore not account for the observed separation and some other source for the magnetic forces must be sought.

In a recent article Pauli[2] showed that it is possible to classify the multiplet levels by assigning four quantum numbers, i.e., four degrees of freedom, to each electron and that this assignment was in general agreement with our knowledge of such levels and his R quantum number is always $\pm 1/2$. In view of the above it seems possible to assume that the new degree of freedom is due to the fact that the electron possesses an intrinsic magnetic field. We do not need to restrict ourselves to any particular model of an

electron which would make it magnetic. Indeed relativity treatment of any such model would require very special considerations. We will assume here that the electron is a charged magnetic doublet of the magnetic moment of one-half a Bohr magneton. For the present considerations we do not need to know the angular momentum though the magnitude of the magneto-mechanical effect suggests that its value is equal to $h/2\pi$.

In our calculations we will neglect the perturbing effect of other electrons in the atom, that is to say our calculation will be strictly applicable only to the case of an electron moving in an inverse square field. The calculation will apply very closely for the X-ray levels of the atom, however, and by using the theory of Landé in regard to the contribution to the energy by the inner and outer parts of a penetrating orbit, it can be applied to the optical levels.

In securing the mutual energy of the system of a magnetic electron moving in the positive field of the nucleus, we can start from the electromagnetic equation that the magnetic force exerted on a unit magnetic pole moving in the electric field is

$$\mathbf{F} = -\frac{\mathbf{v} \times \mathbf{d}}{c},$$

where $\mathbf{v}$ is the vector velocity of the magnetic pole and $\mathbf{d}$ the electric intensity at the point. This force would cause a precession of the rotating electron in the same way as an external magnetic field causes a precession of the electron orbit of the hydrogen atom and the additional energy would be

$$\Delta E = \frac{e\mathbf{H}}{4mc} p_e \cos \alpha$$

$$= -\frac{p_e e}{4mc} \frac{|\mathbf{v} \times \mathbf{d}|}{c} \cos \alpha,$$

where p_e is the mechanical angular momentum of the electron and α is the angle between p_e and the angular momentum of the electron moving about the nucleus. The factor $^1/_4$ is used in place of $^1/_2$ since we are assuming that the mechanical moment is twice the magnetic moment of the electron, both being measured in quantum units. We assume that the quantum number for the rotation of the electron is 1. The same result can be secured if we assume that the rotational quantum number of the electron is $^1/_2$ and that its magnetic moment is $^1/_2$ a Bohr magneton.

In the formula for ΔE we can substitute for $\mathbf{v}$ and $\mathbf{d}$

$$\mathbf{v} = r\varphi\varphi_1 + r\tau_1$$

$$\mathbf{d} = -\frac{Ze}{r^2}\mathbf{r}_1.$$

Then

$$\Delta E = \frac{p_e Z_{e^2}}{4mc^2} \frac{r^2\varphi}{r^3} \cos\alpha$$

$$= \frac{p_e Z_{e^2}}{4mc^2} \frac{p_\varphi}{r^3} \cos\alpha,$$

p_φ being the angular momentum of the electron about the nucleus. We regard this as a perturbing function and take the mean value over the unperturbed motion, i.e., the orbit of a non-magnetic electron moving in the field of a positive nucleus.

$$\frac{\overline{1}}{r^3} = \frac{1}{b^3} = \frac{64\pi^6 e^6 m^3 Z^3}{n^3 k^3 h^6}$$

where b is the semi major axis of the ellipse and n and k are the principal and subordinate quantum numbers. The final form for the energy is

$$\Delta W = \overline{\Delta E} = p_e \frac{8\pi^5 e^8 Z^4 m}{c^2 h^5 n^3 k^2} \cos\alpha.$$

If we set $p_e = \frac{h}{2\pi}$ and $\cos\alpha = \pm 1$,

$$\Delta W = \pm \frac{4\pi^4 e^8 Z^4 m}{c^2 h^4 n^3 k^2},$$

the positive sign being used when p_e and p_φ have the same direction and the negative sign when they are oppositely directed. This formula is the same as that derived by Landé except that the separation of the levels is now proportional to the fourth power of Z instead of the third power.

If following Landé,[9] we write this formula

$$\Delta W = \pm \frac{4\pi^4 e^8 Z_i^4 m}{c^2 h^4 n_i^3 k^2}$$

for the penetrating optical levels, where Z_i is the effective atomic number and n_i the effective total quantum number for the inner part of the orbit and then multiply by the factor

$$\frac{t_i}{t_a} = \frac{n_i^3 Z_a^2}{n_a^3 Z_i^2},$$

where t_i and t_a are the time spent in the inner and outer parts of the path, respectively, and Z_a and n_a are the effective atomic numbers and effective total quantum number, respectively, for the outer part of the orbit, we secure the formula

$$\Delta W = \pm \frac{4\pi^4 e^8 m Z_i^2 Z_a^2}{c^2 h^4 n_a^3 k^2}.$$

This differs from Landé's formula only in having Z_i in the second power instead of the first and as he has shown this agrees approximately with the experimental data on the separation of doublet and triplet levels.

We must choose for the azimuthal quantum number of an s orbit of silver the value $^1/_2$ since the experiments of Stern and Gerlach show that this atom has a total magnetic moment of one Bohr magneton which on our theory is contributed half by the electron rotating on itself and half by its rotation around the nucleus. We must, therefore, choose half quantum numbers for the azimuthal quantum numbers.

Our formula differs from Sommerfeld's only by a small difference then between his k_1k_2 and our k^2. Our k is the mean of his k_1 and k_2 so that $k^2 = k_1k_2 + {}^1/_4$. This difference is probably unimportant since it amounts to only 1 part in 9 for the 2 quantum orbits (the **L** shell relativity doublets) and is smaller for the doublets having higher values of the total quantum number. This seems to be a possible rational explanation of the so-called relativity doublet levels, both in the X-ray and optical region.

We see that the additional energy is positive when the angular momentum of the electron rotating on itself has the same sign as its angular momentum about the nucleus so that the objections of Breit to the original theory of Heisenberg do not apply to this theory; and so far as we can see Heisenberg's theory of the anomalous Zeeman effect for the doublet type of spectrum follows without any change whatever. It is also found that a partial explanation for the anomalous Zeeman effect and multiplicities of spectra from atoms containing more than one electron can be secured, but a complete dynamical description would require a more extensive study. The case of the S level of the alkaline earths is of considerable interest. Each electron must be characterized by $n_k = n_{1/2}$, since the removal of either leaves the ion in the normal state and as mentioned above it is necessary to assign these quantum numbers in order to meet the experiments of Stern and Gerlach.[19] Since the magnetic moment of the normal alkaline earth atom should be zero this means that the two electrons must be oriented in opposite directions. And according to the general scheme of Stoner,[11] this would lead us to conclude that Sommerfeld's[8] model for the helium atom should be correct. It is interesting that this model, though dynamically unstable, does give the correct ionizing potential for helium.

Also on the basis of the present theory we would expect the spectrum of helium to consist of a singlet and triplet series of lines just as in the case of the alkaline earths. It seems that the decision of the question as to whether the ortho helium spectrum is a doublet or triplet spectrum might be decisive.

One serious objection to half quantum numbers has been that above atomic number 68 the electron moving in a $^1/_2$ orbit would collide with

the nucleus if relativity mechanics holds. While a magnetic force acts on the electron moving about a positive charge, an electric force is exerted on the nucleus due to the rotation of a magnet about it. This force is equal to $-\frac{\mathbf{v} \times \mathbf{h}}{c}$, where $\mathbf{v}$ is the velocity of the magnet and $\mathbf{h}$ is the magnetic intensity at the nucleus. This force is in the opposite direction to the electrostatic force of attraction of a nucleus and electron and tends to force them apart if the p_a and p_φ have the same direction. It may be that this force will prevent the collision of nucleus and electron.

We have found it impossible to account for the fine structure of the hydrogen and helium lines by assuming a simple superposition of the relativity change in energy according to Sommerfeld's formula and the magnetic changes in energy according to the above formula, if half quantum numbers are used for the azimuthal quantum numbers. However, it appears that the proper form to apply the relativity theory to this case is possibly not that given above. Furthermore it does not seem impossible that the proper form of relativity mechanics suitable to this electron may give the correct fine structure of the hydrogen and helium lines.

We wish to state that in spite of the rather remarkable results secured for the relativity doublets in the optical and X-ray regions, we appreciate fully the very important fault of the theory, that in its present form it cannot account for the fine structure of the hydrogen and helium lines which is so beautifully explained on the basis of the Lorentz electron. Until this can be done we regard the theory as a purely tentative suggestion.

Note added in proof: Since this paper was written, Dr. W. H. McCurdy and Dr. W. A. MacNair[12] of the Johns Hopkins Physics Department have separated the ortho helium lines and find no evidence for a triplet structure.

[1] G. E. Uhlenbeck and S. Goudsmit, *Naturwissenschaften*, **47**, 953 (1925).
[2] W. Pauli, Jr., *Zs. Physik*, **31**, 765 (1925).
[3] A. L. Parsons, *Smithsonian Inst. Publ.*, **65**, No. 11 (1915).
[4] W. Heisenberg, *Zs. Physik*, **8**, 273 (1922).
[5] G. Breit, *Nature*, **112**, 396 (1923).
[6] R. A. Millikan and I. S. Bowen, *Physic. Rev.*, **23**, 1 (1924).
[7] A. Landé, *Zs. Physik*, **25**, 46 (1924).
[8] A. Sommerfeld, *Atomic Structure and Spectral Lines*, Methuen, **1923**.
[9] A. Landé, *Zs. Physik*, **24**, 88 (1924).
[10] O. Stern and W. Gerlach, *Ibid.*, **8**, 110 (1922); **9**, 349 (1922).
[11] E. C. Stoner, *Phil. Mag.*, **48**, 719 (1924).
[12] W. H. McCurdy and W. A. MacNair, Unpublished note, *Nature*.

THEORY OF THE SPECIFIC HEAT OF ELECTROLYTES*

BY F. ZWICKY†

NORMAN BRIDGE LABORATORY OF PHYSICS, CALIFORNIA INSTITUTE OF TECHNOLOGY

Communicated January 4, 1926

No molecular theory of the specific heat of liquids has been deduced till now. The considerations given in this paper may be regarded as a first attempt to solve that problem. Most molecular theories of solutions regard the solvent as a continuum, treating only the solute as composed of molecules. In this theory this simplification is made wherever possible. For the treatment of certain phenomena, however, it is too great an idealization, and it is necessary to consider the molecular character of the solvent also. As in the theories of other properties of solutions, only differences between the specific heat of the solution and the pure solvent are calculated. The application of the principles to be deduced, to any dilute solution or liquid mixture does not seem to offer any serious difficulties. It seems possible, that by proceeding logically in this direction we may find the key to the explanation of the specific heat of liquids on the basis of molecular models.

We first examine a few cases of un-ionized solutions, drawing certain conclusions therefrom, and then proceed to a general consideration of the so-called strong electrolytes. (The solvent is always supposed to be water.) As shown farther on, the results thus obtained are found to be in very good agreement with the experimental data.

Electrolytes are in general distinguished from non-electrolytes by the following characteristics. If n mols of a substance are dissolved in n_o mols water, so that $n + n_o = 1$, then we have this relation for non-electrolytes

$$C > n_o C_o \tag{1}$$

whereas for electrolytes we find in general

$$C < n_o C_o \tag{2}$$

where C is the heat capacity of one mol solution and C_o that of one mol water. In case (1) for many substances with big molecules an approximate empirical relation can be found,

$$C \sim n_o C_o + n C_S \tag{3}$$

C_S being the molar heat of the solute in the solid state. The meaning of (3) is that the degrees of freedom of a molecule are nearly the same in both the solid state and in solution. Evidently the relation (3) will be expected to hold better for big molecules than for small ones. Table 1 gives some examples of non-ionized aqueous solutions.

TABLE 1

	GLUCOSE $C_6H_{12}O_6$	LACTOSE $C_{12}H_{22}O_{11}$	GLYCERINE $C_3H_8O_3$	TARTARIC ACID $C_4H_6O_6$	UREA CON_2H_4
$\bar{C} = \frac{C - C_o n_o}{n}$	55.2	104.5	56.0	67	24.8
C_o	51.7	98.5	53.0 (for liquid state)	62	19.3

These figures indicate that non-electrolytes behave normally; in other words that the interaction between the molecules of solute and solvent does not affect the specific heat appreciably. The specific heat of these solutions seems to be nearly additively composed of the specific heats of the solvent and solute in the pure states.

As indicated by (2) electrolytes behave quite differently. The interaction between the ions and the solvent (water) cannot be neglected. On the contrary, this effect is predominant. The factors which must therefore be considered in calculating the specific heat of electrolytes are the following:

1. Degrees of freedom of the molecules of the solute.
2. Change of the specific heat of water due to electric polarization in the field of the ions.
3. It is known that the H_2O-molecules act as rigid electric dipoles.[1] Now such a dipole in the inhomogeneous field of an ion will be attracted towards it. The result of this attraction is a high pressure around the ion. As the specific heat of water is very appreciably changed by pressure (see Fig. 2) this effect will play a great rôle.
4. Hydration of the ions.
5. Ion atmosphere. It has been shown by Milner[2] and Debye[3] that according to the Boltzmann principle the ions arrange themselves in such a way that around every positive ion is built up a negative space charge and vice versa. As the temperature is raised, these ion atmospheres are destroyed and the energy required by this process furnishes a positive member in the expression for the specific heat of the solution. But it may be shown that this effect can be neglected for the first approximation because it is proportional to $n^{3/2}$ whereas the other effects are proportional to n.

To have a definite basis for the discussion we consider a solution of n mols KCl in n_o mols water. If instead of the K^+ and Cl^- ions we dealt with neutral particles of the same shape we could write

$$C = n_o C_o + \lambda n \, R/2 \qquad (4)$$

where R is the gas constant and λ the number of degrees of freedom for one molecule in solution. In the case of KCl λ would be 12, if we ascribe $^1/_2 kT$ potential energy to each degree of freedom. If F_o is the free energy of such an ideal solution we find the same function F for the real solution

by substracting from F_o certain amounts of energy A_i. These are determined by considering the following isothermal processes. First we destroy the ion atmospheres by going over to a random distribution of the ions. The energy necessary for this process we call A_1. This is negligible, as mentioned before, and will not be considered further. Then we discharge the ions, keeping their state of hydration, as well as the internal pressure of the water, constant. The energy required for this is A_2. Then we must reduce the internal pressure to one atmosphere, put-

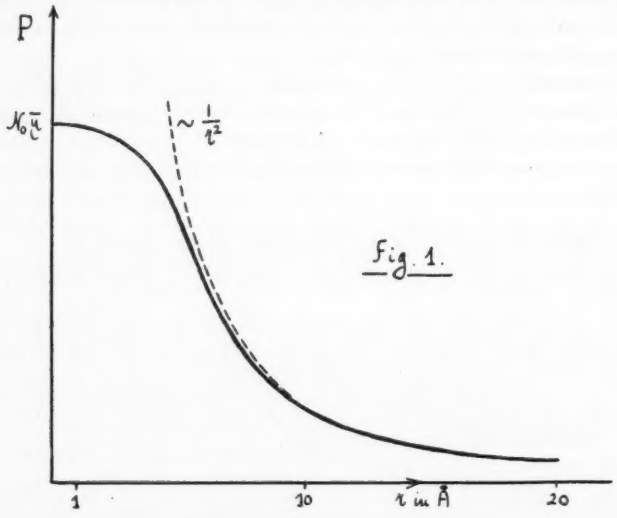


Fig. 1.

ting in the energy A_3. Finally we melt the hydrated complexes with an energy A_4. Then the following relation holds.

$$F = F_o - (A_2 + A_3 + A_4)$$

and the specific heat at constant volume

$$C_v = -T\left(\frac{\partial^2 F}{\partial T^2}\right)_v = -T\left(\frac{\partial^2 F_o}{\partial T^2}\right)_v + T\frac{\partial^2}{\partial T^2}\sum_{i=2}^{4} A_i = \tag{5}$$

$$n_o C_o + n\left[\lambda\frac{R}{2} + C_2 + C_3 + C_4\right]$$

where the C_i are defined by the equation

$$nC_i = +T\left(\frac{\partial^2 A_i}{\partial T^2}\right)_v.$$

In determining the different functions A_i we proceed in the following

way. The assumption of a linear relation between P, the electric polarization of the water, and the electric field would give

$$A_2 = -2nN\frac{\epsilon^2}{2aD} \tag{6}$$

where D is the dielectric constant, a the ion radius, $\epsilon = 4.77 \cdot 10^{-10}$ e.s.u., and $N = 6.06 \cdot 10^{23}$ the Avogadro number. But (6) cannot be quite exact because the very high field in the vicinity of the ion will produce a saturation effect so that P will approach a limiting value $N_0\bar{\mu}$, where N_0 is the

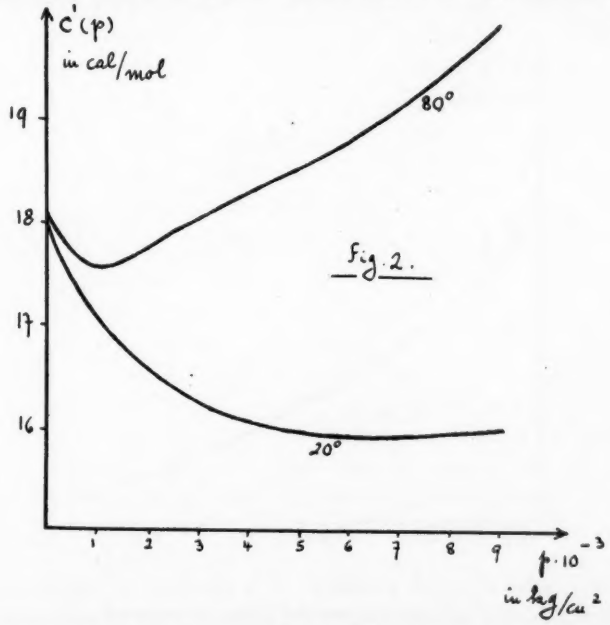


Fig. 2.

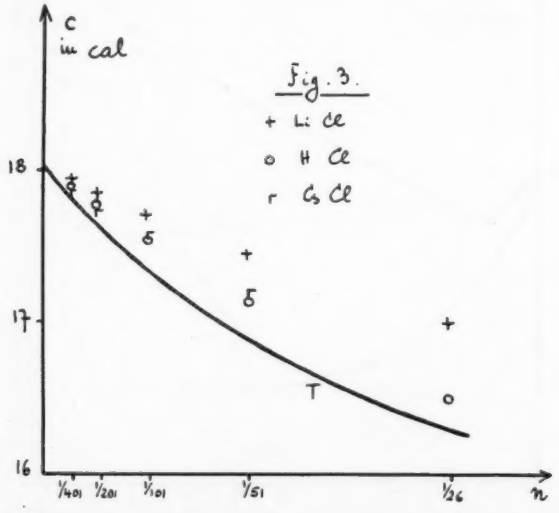
number of watermolecules per cm.^{3} and $\bar{\mu}$ the average of the dipole moment taken over all the molecules, simple and polymerized. It can be shown by considerations analogous to those used in the theory of ferromagnetism (concerning the "molecular field") that there is an upper limit for $\bar{\mu}$ namely $\bar{\mu} = 9 \cdot 10^{-19}$ e.s.u. But $\bar{\mu}$ can be restricted even more as will be shown in another paper which deals with the theory of compressibility, of the coefficient of thermal expansion, and of the change of volume on solution. There it will be found that $\bar{\mu} = 5 \cdot 10^{-19}$ e.s.u. Now P must be equal to $N_0\bar{\mu}$ near the ion and equal $\frac{\epsilon}{r^2}\frac{D-1}{4\pi D}$ far away. The approximation for intermediate values of r will be given by a Langevin function. The resulting curve for $P(r)$ is shown in figure 1.

Calculation of A_2 on the basis of this curve furnishes $C_2 = -10$ cal. per mol. dissolved substance.

In determining the pressure effect (3), it is convenient to calculate C_3 directly. It is easy to show that

$$nC_3 = \int [C'_{(p)} - C_0] \frac{\rho}{M} dv$$

where $C'_{(p)}$ = molar heat of water under pressure p, and ρ is the density of water. The integration is to be extended over all the water contained in the solution. $(C' - C_0)$ is known from Bridgman's measurements[4] (see Fig. 2).



For very dilute solutions we find $C_3 = -119$ cal. At higher concentrations, however, the overlapping of the fields of force of neighboring ions

TABLE 2

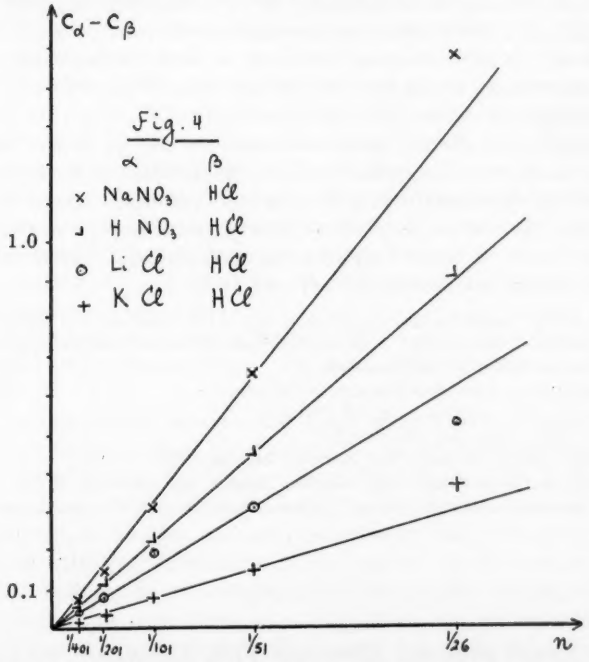
r in Å	1	2	3	4	4.7	5	5.65
p in kg./cm.²	76500	16640	4890	1554	383	124	0
$[C' - C_0]\frac{\rho}{M}$	...	0.100	0.112	0.0692	0.0250	0.00917	0
$r^2[C' - C_0]\frac{\rho}{M}$	...	0.400	1.01	1.11	0.546	0.229	0

must be taken into account, so that the decrease in the specific heat of the solvent produced by pressure will be less than stated above. It will be approximately

$$C_3 = 2N \int_a^{\bar{R}} [C' - C_0] \frac{\rho}{M} 4\pi r^2 dr \qquad (7)$$

where $\bar{R}$ is the effective radius of action for an ion. Table 2 gives the figures necessary for graphical computation of the integral (7) in the case $n = 0.02$ and $\bar{R} = 5.65$ Å.

Combining our results, we find that $C - nC_4 = n_0C_0 + n\left(\lambda\frac{R}{2} + C_2 + C_3\right)$ has the values plotted in figure 3. There are added also some measurements concerning solutions similar to KCl. The curves for NaCl, KCl and RbCl lie between those of HCl and LiCl.[5]



The order of magnitude and also the general shape of the curve is correct. The systematic differences must be ascribed to different hydrations of the ions. Their character is well shown in figure 4, where the difference, $C_\alpha - C_\beta = n(C_{4\alpha} - C_{4\beta})$ for two similar electrolytes α and β, is plotted. As the curves in figure 4 are linear, the effect contributed by the hydration of one ion is constant in this range of concentrations.

Figure 3 shows a somewhat greater curvature for the theoretical curve than for the experimental ones. This is probably because in computing C_3 we have not taken into account the fact that ions of opposite sign are on the average nearer together than charges of the same sign.

The variation of the specific heat of an electrolyte with the valence of

the ions is also in agreement with the theory. A very good check for the latter is found in calculating the variation of the specific heat with temperature. Again C_3 turns out to be the most important factor. Considering the curves in figure 2, it can be shown that C_3 increases very slowly with the temperature from 10° to about 60°C. A sudden change from negative to positive values occurs, however, in the region between 60° and 80°C. The result is, that the true specific heat shows small variation from 10° to 60° and increases very rapidly afterwards. A preliminary check of this strange behavior being predicted by the theory is possible in the case of KCl and NaCl where the average specific heat from 15° to 50° and from 15° to 80° is known. However, it would be desirable to have exact measurements of the true specific heat from 0° to 100° so as to be able to compare a whole curve with the theory.

It is possible to extend these considerations and to deduce the free energy of an electrolyte in terms of volume, temperature and concentration for a relatively dilute solution (up to some mols. per liter). It can be shown furthermore that the pressure effect (3) will probably play an important rôle in the theory of osmotic phenomena, necessitating a correction to the treatment of this problem by Milner and Debye.

* In the present paper we give a general outline of the considerations underlying our theory and some of the results. A detailed account of the intermediate mathematical steps will be published in another place.

† INTERNATIONAL RESEARCH FELLOW IN PHYSICS.

[1] P. Debye, *Phys. Zeitschr.*, **13**, 97 (1912).

[2] Milner, *Phil. Mag.*, **23**, 551 (1912).

[3] P. Debye and E. Hückel, *Phys. Zeitschr.*, **24**, 185 (1923).

[4] P. Bridgman, *Proc. Amer. Acad. Arts and Sciences*, **48**, 310 (1912–13).

[5] T. W. Richards and A. W. Rowe, *J. Amer. Chem. Soc.*, **43**, 770 (1921) and **44**, 684 (1922).

THE SPECTRUM OF NEON IN THE EXTREME ULTRA-VIOLET

BY T. LYMAN AND F. A. SAUNDERS

JEFFERSON PHYSICAL LABORATORY, HARVARD UNIVERSITY

Communicated January 8, 1926

The study of the spectrum of helium in the extreme ultra-violet recently carried through by one of us[1] brought to light a persistent pair of lines apparently due to an impurity, and, as the most reasonable assumption was that this impurity was neon, the investigation of the spectrum of this gas was undertaken. A preliminary account of this work[2] was published

in 1925. G. Hertz has recently announced[3] his independent discovery of the two strong neon lines.

The vacuum grating spectroscope with which this investigation was carried on has been fully described.[1] The grating, of one meter radius, and 15,000 lines to the inch, was ruled by Professor Wood at Johns Hopkins University in the spring of 1924. False spectra are almost completely absent, but it is interesting to note that faint ghosts are present which may be described as of the Rowland type, though their distance from the parent line is unusually large—4 A.U. at λ 740. On the whole, however, the behavior of the instrument has been very satisfactory; the spectra on one side are strong both in the first and second order and the resolving power is excellent. The grating has now been in continuous use for more than a year, yet no diminution in the intensity of its extreme ultra-violet spectra has been observed. This points to a freedom from tarnish which in turn may be attributed to the design of the spectroscope.

In all the experiments a discharge tube of the internal capillary type was employed, the material was quartz except in a few cases where pyrex glass was used. The capillary had a bore of about 3 mm.; its end was usually distant one cm. from the slit. In the standard exposures the electrodes were of tungsten.

The direct current which was usually used to excite the tube had values between 15 and 20 milliamperes. The slit varied between 0.02 to 0.03 mm. in width.

The crude gas which served as the source of the neon used in this investigation was obtained from the Laboratory of the Air Reduction Co. at Jersey City about four years ago. Our colleague, Professor H. N. Davis, to whose kindness we owe this material, has informed us that the method of preparation of this gas was such as to render the presence of krypton and xenon extremely improbable. The gas was purified for us by Dr. H. M. Chadwell during the summer of 1921, by standard chemical methods. The separation of neon from other rare gases was not attempted. It appeared that the original material contained about fifty per cent of neon. Argon and helium were both present. Krypton and xenon could not be detected by spectroscopic examination in the visible part of the spectrum.

In the source of an attempt to identify lines of doubtful origin which appeared in the spectrum of neon, the spectra of argon, oxygen and nitrogen have also been investigated. The first gas was obtained from the Western Electric Co. It was purified by chemical means. The second was received in a state of great purity from our colleague, Professor G. P. Baxter; and the third, also a very pure specimen, we owe to the kindness of Professor Keyes of the Massachusetts Institute of Technology.

The chief results of our investigation are shown in the accompanying plate.

(*a*) shows a heavily exposed neon spectrum, one hour, direct current discharge; (*b*) shows a shorter exposure, one-half hour, and (*c*) is a spectrum of helium in the same region, which contained neon as an impurity.

The photographs show a pair of lines (λλ 743.73 and 735.95) of relatively very great intensity, and a number of others that lie between that wave-length and the helium line at λ 584. In (*a*) the plate has become solarized at the centers of the lines producing a misleading impression of a reversal, and preventing the lines from assuming their true proportions. In (*b*) the exposure was stopped before this had occurred.

Table 1 shows the wave-lengths of the lines in this region which appear to belong to neon in its normal state (Ne I). All the lines between λ 630 and λ 590 which show on the plate are of this sort, though on the original photographs a few fainter ones also appear, which we are not prepared to assign definitely to this spectrum. The group near λ 670, and the other lines near λ 730 appear not to be due to neon but to some unknown impurity. In particular they seem not to belong to ionized neon; this spectrum also we have examined, though not exhaustively. It is a rich one, extending with considerable intensity to λ 430 and more faintly even further. The identification of impurities is not an easy task and we are not yet able to establish the origin of all the lines on our neon plates. A particularly troublesome group appears in the region above λ 800, but from the evidence at hand we feel convinced that these are not due to neon in the normal state. A special survey of the spectrum from λ 2500 to λ 750 failed to reveal any Ne I lines in this region, nor have we found any below λ 584. We have obtained the spectrum in tubes of pyrex and of fused quartz, and with electrodes of tungsten, nickel and aluminium, without finding any reason to doubt our identifications.

The wave-lengths of the stronger lines were determined from several plates, using hydrogen series lines as standards against second order images of neon and helium lines. They should be accurate to within 0.2 A.U.

Table 1 shows also the series relationships of the different lines, using Paschen's notation. The fundamental $1p$ level combines only with levels of inner quantum number 1 on Landé's first scheme[4] (or $1^1/_2$ in the later numbering). It therefore should have the inner quantum number 0, as predicted by Goudsmit.[5]

Table 2 shows the values of $1p$ as obtained from each line, by adding to its wave-number the appropriate term as given by Paschen. The variations in the value of $1p$ are small enough to indicate a probable error of less than 0.1 A.U. but systematic errors of larger size may be present. The average value 173,930 points to an ionization potential of 21.47 volts, as already announced by us and by Hertz, in agreement with the directly determined value of 21.5 given by Hertz.

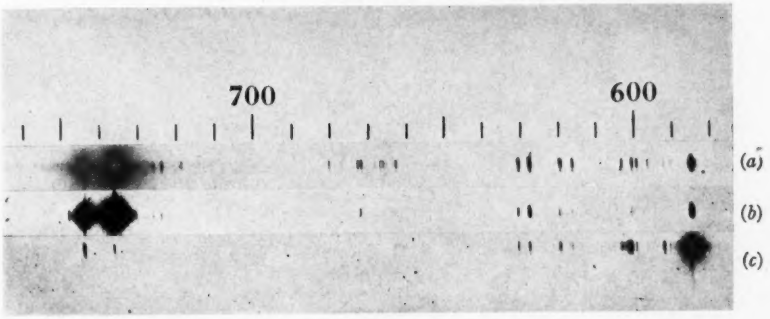


PLATE 1
Enlarged × 10.

TABLE 1

INT.	λ VAC.	ν	SERIES	
9	743.73	134457	$1p$–$1s_4$	
10	735.95	135879	$1p$–$1s_2$	
4	629.75	158793	$1p$–$2s_4$	
5	626.91	159513	$1p$–$2s_2$	
3−	619.04	161540	$1p$–$3d_5$	
4	618.67	161637	$1p$–$3d_2$	
2+	615.62	162438	$1p$–$3s_1'$	
2+	602.74	165909	$1p$–$3s_4$	
3	600.04	166656	$1p$–$3s_2$	
2	598.81	166998	$1p$–$4d_2$, d_5	not resolved
1+	598.89	167816	$1p$–$4s_1'$	
1	591.80	168976	$1p$–$4s_4$	
1	589.93	169512	$1p$–$5d_2$, d_5	not resolved
1−	589.17	169730	$1p$–$4s_2$	
0+	587.17	170308	$1p$–$5s_1'$	
0+	586.41	170529	$1p$–$5s_4$	

TABLE 2

VALUE OF $1p$ FROM OBSERVED WAVE NUMBER BY ADDING APPROPRIATE TERM

TERM ADDED (FROM PASCHEN)		OBSERVED WAVE NO.	$1p$
$1s_4$	39470	134457	173927
$1s_2$	38041	135879	20
$2s_4$	15141	158793	34
$2s_2$	14396*	159513	29
$3d_5$	12405	161540	45
$3d_2$	12293	161637	30
$3s_1'$	11494	162438	32
$3s_4$	8017	165909	26
$3s_2$	7273	166656	29
$4d_2$	6902	166998	27
$4d_5$	6954		
$4s_1'$	6122	167816	38
$4s_4$	4962	168976	38
$5d_2$	4412	169512	38
$5d_5$	4441		
$4s_2$	4202	169730	32
$5s_1'$	3633	170308	41
$5s_4$	3372	170523	01
		Weighted average	173930

* Paschen gives this value in his second neon paper (*Ann. Physik*, **63**, 201, 1920). An earlier, incorrect value of 14506 is found in Paschen-Götze.

Plate 1 (*c*) shows a curious effect. When neon is present in small quantities as an impurity in helium, we find the intensity of λ 735 often very greatly reduced, as though this line were somehow sensitive to conditions which we have not been able to determine. One might suppose that there

is an absorption band, on the edge of which this line lies; but we have definite evidence against this supposition. Photographs of the disruptive discharge in helium show strong lines, due to impurities, throughout the whole region where such an absorption band would have to lie. There seems to be no obvious explanation of the behavior of this line under these circumstances. It is interesting to note that when argon is present as an impurity in neon, its line λ 1048.26 (which corresponds to λ 735 in neon) shows no such variations in intensity. The next line in the same neon series, namely $1p$–$2s_2$ at λ 626.9 is of normal intensity, as may be seen in the plate (*a*).

Mr. H. W. Leighton has carried out much of the experimental manipulation necessary to this investigation. It is a pleasure to acknowledge the debt which we owe to his patience and skill.

[1] T. Lyman, *Astroph. J.*, **60**, 1 (1924).
[2] *Physic. Rev.*, **25**, 886 (1925).
[3] *Zs. Physik.*, **32**, 933 (1925).
[4] *Ibid.*, **17**, 292 (1923).
[5] *Ibid.*, **32**, 111 (1925).

THE FREE PATH OF PROTONS IN HELIUM

By Arthur J. Dempster

Ryerson Physical Laboratory, University of Chicago

Communicated December 23, 1925

In a former paper[1] experiments were described with protons that had been accelerated through 900 volts' potential difference. The particles were found to have a much longer free path in hydrogen than would be expected from Rüchardt's observation with faster rays (13,000 to 16,000 volts). These experiments have since been extended to pressures five times as great as previously used, with similar results.

Experiments with the same apparatus have been made with protons in helium. The source was an anode of lithium that had been allowed to oxidize, and probably form the hydroxide. On bombardment with electrons it served as a source of protons, although other ions were also given off, hydrogen molecules, water vapor and nitrogen molecules. The rays were accelerated through a definite potential and deflected into a semicircular path by a magnetic field. They then fell through a slit onto an electrode where their charge was measured by an electrometer. As described in former papers,[2] the strength of the magnetic field and accelerating voltage serves to determine the nature of the charged particles

falling on the electrode. The length of the path completed by the charged particles was 17 cms.

Helium gas was admitted at various pressures, and the different rays were observed by varying the magnetic field so as to bring them past the slit in front of the collecting electrode. At comparatively low pressures all the ions present except the hydrogen atoms were much reduced in intensity and disappeared at a pressure that had very little influence on the intensity of the bundle of protons. As the pressure increased, this bundle was broadened, but, with 930 volts accelerating potential, protons were observed as a definite group of rays with a pressure of 0.53 mm. of mercury. At this pressure the mean free path for a rapidly moving particle, assuming the kinetic gas theory value for the diameter of the helium atom, is 1.4 mms., so that the protons must have passed through more than 120 atoms without having their velocity or direction radically altered. Protons were observed with an applied accelerating potential difference as low as 14 volts. The slower rays were weaker and were not observed at such high pressures. This is probably to be accounted for by greater scattering rather than by any difference in their ability to penetrate the atoms. Rays obtained with 14 volts were observed with pressures up to 0.03 mm., at which pressure they would make seven collisions in the path completed.

Observations were made to test whether the protons lost energy in passing through the atoms, by noticing whether the magnetic field required to keep the rays in the same path remained the same when the pressure was increased. No certain indications of any loss of velocity were obtained.

We are thus led to the conclusion that protons with the velocities used in these experiments will pass freely through atoms of helium, with only slight changes in their velocity or direction.

Recent experiments by Henderson[3] and Rutherford[4] have shown that the alpha particle alternates rapidly between a doubly and singly charged condition. The free path for the capture of an electron decreases rapidly as the velocity decreases and approximates the kinetic theory free path for the lowest velocity. Rüchardt[5] has found that the free path for neutralization of protons accelerated by 13,000 to 16,000 volts, is slightly less than the free path between collisions on the kinetic theory. Theories have been proposed by Rüchardt[6] and Fowler[7] for the rate at which these rapidly moving charges should capture electrons. Both theories suppose electrons to be present that are free to combine with the moving nucleus at every collision. The slower the velocity of the nucleus, the more likely is the capture of an electron to take place.

We might explain the absence of neutralization observed in the experiments with slow protons as due to the inability of the proton of the speed

used to produce free electrons, or to ionize the helium atom. J. J. Thomson[8] has pointed out that the maximum energy transferred to an electron of mass m by the direct impact of a mass M is only a small fraction $4m/M$ of the energy of the larger mass, provided the laws of mechanics apply. We should thus expect no ionization in helium with protons that have fallen through less than 11,270 volts. The application of simple mechanical considerations to this problem is quite doubtful, however, since the development of our ideas of excited states in the atom and the transformation of this atomic energy into kinetic energy of electron or atom. A discussion of the ionization potential of positive ions from this point of view has been given by Franck.[9]

In order to account for the ionization observed by many experimenters when gases are bombarded by ions from hot salts, Thomson suggests that positive particles may be able to capture an electron from an atom and so leave a gas ion, if it is moving with sufficient velocity to pass through the neutral atom. Analogous considerations have been advanced by Franck[10] in discussing the rôle of electron affinity in the change of sign in canal rays, and also in the paper just referred to. The failure of the proton to capture an electron may be associated on this point of view with the fact that the ionization potential of helium is 24.5 volts, while the energy of formation of a neutral hydrogen atom is only 13.6 volts, so that there is no tendency for an electron to leave the helium atom in order to join with the hydrogen nucleus. Ionization tracks in air due to rapidly moving protons have been observed by Blackett[11] in the Wilson drop track apparatus, and we should expect that at some speed protons would acquire the ability of disrupting the helium atom. The ratio of the velocity of the proton to the velocities of the electrons in their orbits may be the determining element in the interaction. The protons used in these experiments have much slower velocities than the electrons in the atom, so that they may be thought of as causing only a slow modification of the electronic orbits during their passage through the atom, without producing any permanent disturbance or ionization.

The author's thanks are due to Mr. G. E. Read for assistance in making the observations.

[1] A. J. Dempster, *Proc. Nat. Acad. Sci.*, **11**, 552–554 (1925).
[2] A. J. Dempster, *Physic. Rev.*, **20**, 631–638 (1922).
[3] Henderson, *Proc. Royal Society, London*, **102**, 496–505 (1922).
[4] E. Rutherford, *Phil. Mag.*, **47**, 277–303 (1924).
[5] E. Rüchardt, *Ann. Physik*, **73**, 230–236 (1924); **71**, 380–423 (1923).
[6] E. Rüchardt, *Zeit. Physik*, **15**, 164–171 (1923).
[7] R. H. Fowler, *Phil. Mag.*, **47**, 416–430 (1924).
[8] J. J. Thomson, *Rays of Positive Electricity*, 2nd Edition, p. 56–59.
[9] J. Franck, *Zeit. Physik*, **25**, 312–316 (1924).
[10] J. Franck, *Physik. Zeit.*, **14**, 623–624 (1913).
[11] P. M. S. Blackett, *Proc. Royal Society, London*, **107**, 349–360 (1925).

SUBSIDENCE RATE OF REEF-ENCIRCLED ISLANDS

BY W. M. DAVIS

HARVARD UNIVERSITY

Communicated December 27, 1925

The subsidence of reef-encircled islands and of atoll foundations is indicated, as I have elsewhere shown in greater or less detail, by various lines of evidence; for example, by the disappearance of large volumes of detritus from maturely dissected volcanic islands,[1] like Raiatea and Huaheine in the Society group; by the occurrence of drownd-valley embayments with estimated rock-bottom depth of 600 or 800 feet at the mouth, as in Tahiti and New Caledonia;[2] by the seaward transition of mountainous coasts into lowering archipelagoes, as in northeastern Australia;[3] by the considerable breadth of lagoons enclosed by barrier reefs based on volcanic islands of simple structure,[1] as around Tahaa in the Society group and Kandavu and other islands in the Fiji group; by the small size of mountain-top islets in the large lagoons of almost-atolls,[4] like those of Truk in the western Carolines and of Mangareva southeast of the Paumotus; by the unconformable contact of elevated reef limestones with their foundations,[5] as in Vanua Mbalavu and Tuvuthá in eastern Fiji; by the occurrence of occasional atolls in the neighborhood of demonstrably unstable island groups, such as the Solomon and Louisiade groups[6] in the far-western Pacific, as well as in Fiji;[5] and it may be added that such occurrence is of special interest because the origin of atolls is usually inscrutable. Furthermore, indirect evidence for the subsidence of reef-encircled volcanic islands and of atoll foundations in the coral seas of the torrid Pacific is given by the absence of truncated old volcanic islands in the cooler seas of the North and South Pacific.[7]

Additional evidence for the strong subsidence of certain islands is given by the following considerations: Let a volcanic island of initial slope, *AB*, figure 1, suffer mature dissection, so that its spur profiles are reduced to *CDB*, and its stream profiles to *CEB;* and let it at the same time suffer subsidence, so that successive sea levels are indicated by horizontal lines, F^1 to F^5; the subsidence being of such rate compared to degradation that the island top, *C*, shall still have strong slopes, such as characterize the islets of almost-atolls, when it is nearly submerged. Small ratios of bay length to spur radius will prevail during early and late stages of submergence; that is, D^1E^1 will be a small fraction of D^1F^1; and similarly D^5E^5 will be a small fraction of D^5F^5. Maximum ratios will obtain during intermediate stages of submergence; that is, bays lengths, D^2E^2, D^3E^3, D^4E^4, may then be a third or a half of their spur radius, D^2F^2, D^3F^3, D^4F^4. A large ratio of bay length to spur radius therefore indicates that sub-

sidence has submerged from one-third to two-thirds of the original height of the island or, in other words, that the subsidence is then to be measured by from half to double the visible height of the island $F°C$. A strongly embayd and still vigorously mountainous island, like Kandavu in southwestern Fiji or Tahaa is the Society group,[8] must therefore have subsided by some such amount as 1000 or 2000 feet. A barrier reef surrounding a well embayd and therefore well-submerged island and enclosing a lagoon, R^2D^2, R^3D^3, R^4D^4, of somewhat less or more width than the corresponding spur radius, D^2F^2, D^3F^3, D^4F^4, must therefore have been built up from a considerable depth, as Darwin justly inferd. This is a very simple argument, but I believe it has not been hitherto presented.

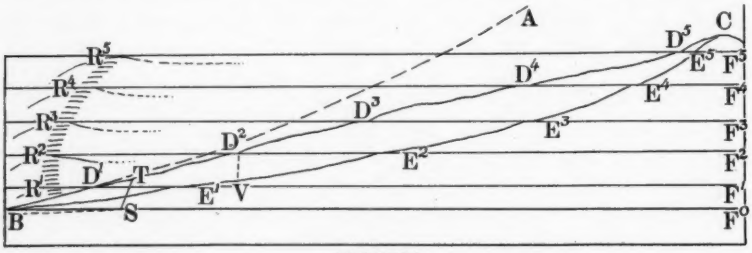


FIGURE 1

It should be noted that the subsidence, $F°F^2$, which an embayd island, CD^2F^2, has sufferd is not fully measured by the rock-bottom depth, D^2V, of the submerged valley at the bay mouth, but by the rock-bottom depth of the valley mouth at B, where the prolongation of the valley profile, CE^2B, meets the prolongation of the spur profile, CD^2B; or if, as is probable, the island was somewhat abraded and clift, BST, in its reefless youth before effectiv subsidence began,[9] the subsidence demanded by such a reef as R^2, would be measured up from the cliff base, S, which may lie significantly below the level of the submerged hanging-valley mouth in the cliff face. An almost-atoll reef, enclosing a broad lagoon, R^5D^5, in which the bay length, D^5E^5, is small even compared to the small spur radius, D^5F^5, would appear to have been built up during a subsidence that may well measure five or ten times the height of the surviving islet, F^5C.

Let it be explicitly noted, however, that the above argument holds good only for islands of fairly uniform resistance thruout their mass. An island of composit structure—for example, an uplifted volcanic island bearing on its flanks a well-developt barrier reef, the lagoon limestones of which heavily envelop but do not completely cover the dissected island summit—may, after certain measures of erosion, re-submergence and new reef upgrowth, have ratios of lagoon breadth and bay length to the subsidence that determined them very different from the ratios above

given. Here, after erosion has degraded the relativly weak limestones to a lowland and has re-excavated the previously embayd and aggraded valleys—at the same time modifying to a much less extent the resistant slopes of the central volcanic island—a slight subsidence will permit the upgrowth of a relativly thin reef, which will nevertheless enclose a broad lagoon, the waters of which will form good-sized bays in the re-excavated central valleys. Vanua Mbalavu in the large lagoon of the Exploring Isles in eastern Fiji appears to exemplify this special case,[5] altho by reason of an eastward increase in the renewd subsidence the lagoon gains the unusual depth and the barrier reef the unusual thickness of 80 or 90 fathoms on the eastern side. The Lesser Antillean island of St. Croix, as described by Vaughan, appears to be on the way to exemplify the same case, but its re-subsidence has as yet caused the submergence of only a narrow part of its degraded limestone area.

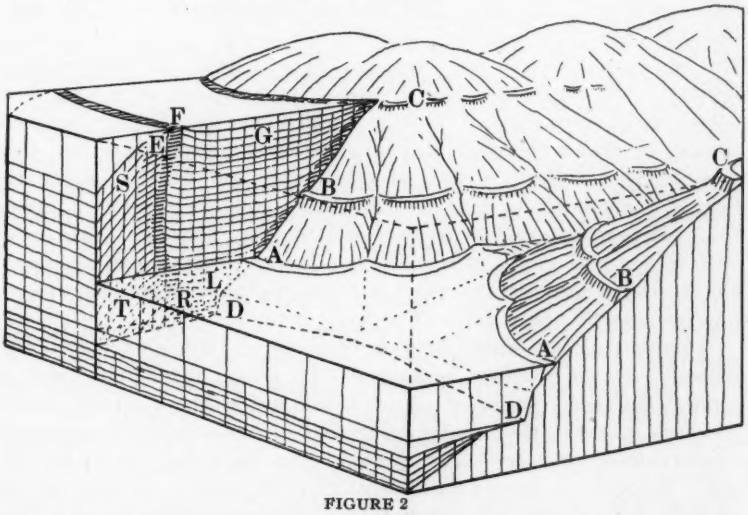


FIGURE 2

The small thickness of certain elevated fringing reefs, such as those described by Guppy in the Solomon islands, has been adduced as evidence that the barrier reefs of the open Pacific are also of small thickness; but the proper interpretation of such elevated fringing reefs shows that their small thickness has no bearing on the presumably greater thickness of well off-set, sea-level barrier reefs, as may be made clear by the following considerations: Elevated fringing reefs, *BB*, *CC*, figure 2, usually lie unconformably on a previously eroded island slope; and Guppy's descriptions show that this is the case with the elevated reefs of the Solomon islands. Inasmuch as the present shore line, *AA*, of the eroded slope is embayd,

the island must have stood higher than now while it was eroded, so that the shore line of that time, *DD*, may now lie several hundred feet below the present shore line, *AA*. Since then the island has subsided at least to *CC*, and after that subsidence it has been elevated so as to place the shore line at *AA*. Whether reef *BB* was formd during a pause in the subsidence and therefore before reef *CC*, or during a pause in the elevation and therefore after reef *CC*, is indeterminate and in the present connection immaterial; but it is pertinent to recall that the production of a fringing reef, such as *CC*, during a pause after a rapid subsidence was explicitly included as a possibility in Darwin's theory;[10] and here that possibility is evidently an actuality.

The absence of continuous reef growth at levels between *DD* and *CC* shows that the subsidence of the island, after the erosion of its slopes with respect to baselevel at *D*, took place so rapidly that it could not be counterbalanced by reef upgrowth. Had the total subsidence by which the shore line was changed from *D* to *C* taken place slowly, a thick barrier reef, *RF*, enclosing heavy lagoon deposits, *LG*, and flankt by exterior talus deposits, *TS*, might have been formd. It is certainly significant that islands which have embayd coasts and which bear elevated fringing reefs of small thickness at various heights above their present shore line, thus demonstrating changes of level at rates faster than that of barrier-reef upgrowth, are found chiefly in the Australasian archipelago, which is known from independent evidence to have sufferd great and rapid changes of level in late geological times; while well developt barrier reefs are found chiefly around islands of the open Pacific where changes of level have therefore presumably taken place at a slower rate.

It appears to be assumed by some objectors to Darwin's theory of upgrowing reefs on subsiding foundations that lagoon aggradation must be significantly slower than reef upgrowth; and hence that, if a thick barrier reef grew up during the strong subsidence of its central island, a deep moat should remain, imperfectly filld, between the diminishing island and the reef. Inasmuch as barrier reefs actually enclose relativly shallow lagoons and not deep moats, it is concluded by these objectors that such reefs cannot have had a great measure of upgrowth. The assumption on which this conclusion is based appears to be unwarranted, as may be made clear by the following argument: Let the assumption be temporarily accepted; that is, let it be believd that reef upgrowth is usually faster than lagoon aggradation; and then let it be postulated that the subsidence of a reef foundation may proceed at three rates; namely, at a slow rate, which may be counterbalanced by lagoon aggradation as well as by reef upgrowth; at a rapid rate which cannot be counterbalanced even by reef upgrowth and at an intermediate rate, which can be counterbalanced by reef upgrowth but not by lagoon aggradation. Around islands that suffer slow

subsidence, shallow reef-enclosed lagoons would be formd; if the subsidence were extra slow, the enclosing reef would be extra broad and the lagoon extra shallow; and this case may be illustrated by the reefs and lagoons of Mbengha and Ovalau in western and central Fiji; if the subsidence were less slow, the reef would be narrower and the lagoon would be deeper; and this case may be exemplified by Budd reef in northeastern Fiji, or by Clipperton atoll in the eastern Pacific. Around islands that suffer rapid subsidence, no continuously upgrowing reefs could be formd, but separate fringing reefs might be built during pauses at different levels, as is illustrated in figure 2 for the Solomon islands. Around islands that suffer subsidence at an intermediate rate, the upgrowing reef would enclose a moat-like lagoon, the depth of which would be a large fraction of the reef thickness. But no examples of such lagoons are known. Hence the assumed inequality between reef upgrowth and lagoon aggradation, above temporarily accepted, must be rejected, and these two processes must be regarded as of not very unlike rates. Only two rates of subsidence need then be postulated, slow and rapid; slow subsidence being of such a rate that it is counterbalanced by lagoon aggradation nearly as well as by reef upgrowth; while rapid subsidence is of such a rate that it cannot be counterbalanced by either of these slow processes. It may be here added that Daly's tables, in which large lagoons are shown to be on the average deeper than small lagoons, suffice to prove that a good share of detritus for lagoon aggradation is supplied by overwash from the outer or growing face of the enclosing reef; but as it occasionally happens that even a small lagoon may have an unusually great depth, as is the case with the above-named little Clipperton atoll in the eastern Pacific, which has well-attested lagoon soundings of over 50 fathoms, it follows that slow subsidence does not always proceed at precisely the same rate.

The rate of island subsidence in the open Pacific is thus shown to be frequently not faster than that of reef upgrowth and lagoon aggradation; but let it be noted that this implies that the rate of reef upgrowth is significantly slower than that of coral upgrowth; for the growing corals, usually limited to a narrow belt, *E*, figure 2, known as the growing face, on the exterior slope of the reef, appear to be the chief supply of material not only for the reef face itself, but also for four other areas; namely, for the upbuilding of the reef flat, *F*, which is usually devoid of growing corals, and which is in some cases much broader than the growing face; for a good part, as above noted, of the aggradation of the lagoon floor, *G*, which is often much broader than the reef flat; for the upbuilding of the gentle slope below the growing face; and for the outbuilding of the steep pitching talus, *S*, which descends to great depths. True, lagoon aggradation is in part accomplisht by locally formd organic deposits; and true again, that a reduction in the amount of growing-face material needed elsewhere

may be effected if the upgrowing reef slants inward toward the central island, for thereby the lagoon will be narrowd and the outbuilding of the talus will be diminisht. But, however great the economies thus effected, it still remains true that reef upgrowth must be slower than coral upgrowth.

The relation of island subsidence and the equivalent reef upgrowth to island degradation is indicated in a general way by the islets of almost-atolls. They are, as above noted, of rounded, mountain-top form, *C*, figure 1, with fairly steep side slopes; hence they must represent the surviving parts of a volcanic cone that has lost about a third, a half or two-thirds of its initial height by erosion, and that has lost most of the rest of its initial height by subsidence. In the case of an almost-atoll reef 12 miles in diameter, the initial height of the much degraded and now nearly submerged island may well have been from 5000 to 7000 feet; hence if the islets of such an almost-atoll are from 300 to 500 feet high, the subsidence may have been 2000 or 3000 feet. In the case of a submerged island and barrier reef, such as the recently discoverd Stewart bank in the China sea[11] appears to be, its depth of about 200 fathoms or 1200 feet must have been brought about by a subsidence at a faster rate than that of reef upgrowth. Two drownd atolls, known as Macclesfield and Tizard banks, in the same sea also appear to have been submerged by a rapid subsidence of smaller amount, after they had been built up in atoll form during a previous subsidence of greater amount.[12]

The relation of island subsidence and reef upgrowth to changes of ocean level in the Glacial period may be briefly considerd. The lowering of the ocean by the withdrawal of water to form continental ice sheets has been calculated by various geologists, with rather widely differing results. According to Daly, the lowering was 30 or 40 fathoms, or from 200 to 250 feet. If the Glacial period included four Glacial epochs, there must have been 8 changes of level, or a total change of from 1600 to 2000 feet. If the entire duration of these changes were about as long as the combined duration of normal ocean level during Interglacial epochs and of lowerd ocean level during the climax of the Glacial epochs, the above total change of 1600 or 2000 feet during the coming and the going of the Glacial epochs would correspond to a change of from 3000 to 4000 feet during the whole Glacial period. This would seem to involve a rate of change in ocean level even more rapid than the rapid changes of island level in the Australasian archipelago during Pliocene and Pleistocene times, and therefore all the more in excess of the average rate of reef upgrowth; and therefore also in excess of the usual rate of island subsidence in the open Pacific. If the 200-foot rise of the ocean to its normal level after the last Glacial epoch occupied 50,000 years, that would demand a rise of ocean level by one foot in 250 years, and this rate would seem to be several times faster than the ordinary mid-Pacific rate of island subsidence and reef upgrowth.

It is perhaps in part because of the rapidity of Postglacial ocean rise as well as because of the unfavorable temperature of ocean water in the marginal belts of the coral seas during the early stages of the rise, that novice Postglacial reefs are now so imperfectly built up to ocean level around the islands in those belts,[13] the Preglacial reefs there having been destroyd by low-level abrasion in the Glacial epochs, as postulated in Daly's Glacial-control theory. On the other hand, it is probably because the strong Preglacial and Interglacial reefs around islands in the true coral seas were not destroyd by low-level abrasion in the Glacial epochs that whatever erosional damage they sufferd during their temporary emergence at times of lowerd ocean level has now been so well repaird in Postglacial time.[14] The ordinary relation of reefs and lagoons in the coral seas, which Darwin explaind by the "renovating agency" of subsidence, may be in good part ascribed to the relatively rapid Postglacial rise of ocean level on reef and lagoon structures which were built up during a slow, intermittent subsidence that was mostly accomplisht during an earlier and much longer time than the Glacial period; these veteran Preglacial reefs having been more or less degraded during successiv Glacial emergences, and finally repaird to their present stalwart form during and after their Postglacial submergence. But reasonably inferd and significant as these oscillations of ocean level and the associated changes of reef form appear to be, it has not been possible, as far as my studies have gone, to demonstrate their occurrence by observable features of the reefs or islands in the true coral seas; they are identified only in the islands and reefs of the marginal belts of the coral seas, and there but imperfectly.[7]

[1] "Subsidence of Reef-encircled Islands," *Bull. Geol. Soc. Amer.*, **29**, 1918, 489–574.

[2] "Les falaises et les récifs coralliens de Tahiti," *Ann. Géogr.*, **27**, 1918, 241–284. "Les côtes et les récifs coralliens de la Nouvelle-Calédonie," *Ibid.*, **34**, 1925, 244–269, 332–359.

[3] "The Great Barrier Reef of Australia," *Amer. J. Sci.*, **44**, 1917, 339–350.

[4] "The Origin of Almost-atolls," *Nature*, **105**, 1920, 292–293.

[5] "Extinguisht and Resurgent Coral Reefs," *Proc. Nat. Acad. Sci.*, **2**, 1916, 466–471. "The Origin of Certain Fiji Atolls," *Ibid.*, **2**, 1916, 471–475. Also, W. G. Foye, "Geological Observations in Fiji," *Proc. Amer. Acad. Arts & Sci.*, **54**, 1918, 1–145.

[6] "The Coral Reefs of the Louisiade Archipelago," *Proc. Nat. Acad. Sci.*, **8**, 1922, 97–151. "The Barrier Reef of Tagula, New Guinea," *Ann. Assoc. Amer. Georgr.*, **12**, 1922, 7–13.

[7] The Marginal Belts of the Coral Seas," *Proc. Nat. Acad. Sci.*, **9**, 1923, 292–296.

[8] "Problems Associated with the Origin of Coral Reefs," *Sci. Monthly*, **2**, 1916, 313–333, 479–501, 557–572.

[9] "Clift Islands in the Coral Seas," *Proc. Nat. Acad. Sci.*, **2**, 1916, 284–288.

[10] "Fringing Reefs of the Philippine Islands," *Ibid.*, **4**, 1918, 197–204.

[11] "The Stewart Bank in the China Sea," *Science*, **62**, 1925, 401–403.

[12] "Coral Reefs and Submarine Banks," *J. Geol.*, **26**, 1918, 198–223, 289–309, 385–411.

[13] "The Formation of the Lesser Antilles," *Proc. Nat. Acad. Sci.*, **10**, 1924, 205–211. See also reference 7, above.

[14] "The Depth of Coral-reef Lagoons," *Proc. Nat. Acad. Sci.*, **9**, 1923, 396–401.

SUMMARY OF RESULTS AND PROOFS CONCERNING FERMAT'S LAST THEOREM

BY H. S. VANDIVER[1]

DEPARTMENT OF PURE MATHEMATICS, UNIVERSITY OF TEXAS

Communicated January 6, 1926

We shall divide the discussion as usual into two cases. If in the relation

$$x^p + y^p + z^p = 0 \tag{1}$$

with x, y and z prime to each other, xyz is prime to the odd prime p, this will be referred to as Case I, and if one of the integers x, y and z is divisible by p, this will be called Case II.

For the most part the proofs of the results given below will be indicated only. Full details are contained in several articles to be presented for publication to the *Transactions of the American Mathematical Society*.

If (1) is satisfied in integers prime to each other and $z \not\equiv 0 \pmod{p}$, then in another paper[2] I gave the relation

$$\prod_{\nu=1}^{k-1} \prod_{r=1}^{[\nu p/k]} (x + \alpha^{[1:r]}y) = \alpha^{-kyq(k)/(x+y)}\omega^p \tag{2}$$

where k is an integer, $1 < k < p$,

$$q(k) = \frac{k^{p-1} - 1}{p}$$

$[S]$ is the greatest integer in S; ω is an integer in the field $\Omega(\alpha)$; $\alpha = e^{2i\pi/p}$; $[1:r]$ is the integer i in the relation $ri \equiv 1 \pmod{p}$.

Let n be a prime $\not\equiv 0$ or $1 \pmod{p}$ and suppose that $xy \not\equiv 0 \pmod{n}$ then $x^{n-1} - y^{n-1} \equiv 0 \pmod{n}$, hence there is some integer in the set $0, 1, \ldots, n-2$, such that

$$x + \beta^a y \equiv 0 \pmod{\mathfrak{p}} \tag{3}$$

where β is a primitive $(n-1)$th root of unity, and $\mathfrak{p}$ is a particular prime ideal divisor of n in the field $\Omega(\alpha\beta)$. If θ is an integer in $\Omega(\alpha\beta)$, θ prime to (p) and $\mathfrak{p}$, then we define the pth power character of θ as $(\theta/\mathfrak{p})$ which, is the power of α appearing in the relation

$$\theta^{\frac{N(\mathfrak{p})-1}{p}} \equiv \left\{\frac{\theta}{\mathfrak{p}}\right\} \pmod{\mathfrak{p}},$$

where $N(\mathfrak{p})$ is the norm of $\mathfrak{p}$.

In (2) set α^m for α and, in the resulting equation take the pth power characters of each member with respect to $\mathfrak{p}$. Noting that

$$x + \alpha^c y = x + \beta^a y + y(\alpha^c - \beta^a),$$

and using (3), we have

$$\left\{\frac{x+\alpha^c y}{\mathfrak{p}}\right\} = \left\{\frac{\alpha^c-\beta^a}{\mathfrak{p}}\right\},$$

and therefore

$$\left.\begin{aligned} &\mu(k-1)D_o m^{p-1} + \Sigma(m[1:r])^{p-2}D_1 \\ &+\ldots+m\Sigma[1:r]D_{p-2} - \frac{mky}{x+y}q(k)I(\alpha) \\ &\equiv 0 \pmod{p}, \end{aligned}\right\} \quad (4)$$

where $\mu = p-1/2$, Σ stands for

$$\sum_{\nu=1}^{k-1}\sum_{r=1}^{[\nu p/k]}, \quad \left\{\frac{\theta}{\mathfrak{p}}\right\} = \alpha^{I(\theta)},$$

$$D_s = \sum_{d=1}^{p-1} d^s I(\alpha^d - \beta^a).$$

Letting m range over the integers $1, 2, \ldots, p-1$, we obtain by elimination

$$\left.\begin{aligned} &D_o \equiv \Sigma\,[1:r]^{p-s}D_{s-1} \equiv 0 \pmod{p}, \\ &\qquad s = 2, 3, \ldots, p-2; \\ &\Sigma[1:r]D_{p-2} - \frac{kyq(k)}{x+y}I(\alpha) \equiv 0 \pmod{p}. \end{aligned}\right\} \quad (5)$$

We also have, if one of the integers x, y, z, say y, is divisible by p,

$$\prod_h (z + \alpha^{[1:h]}x)^{p-1} = p^{\mu}\alpha^{\mu\Sigma[1:h]}\omega_1^p \quad (6)$$

where h ranges over the positive integers $< p$, such that $h + |rh| > p$, $|rh|$ representing the least positive residue of rh, modulo p, and ω is an integer in $\Omega(\alpha)$. Taking pth power characters of both sides of this with respect to $\mathfrak{p}$, and proceeding as in the preceding treatment of (2) we obtain

$$\left.\begin{aligned} &D'_s \equiv \Sigma[1:h]^{p-s}D'_{s-1} \equiv 0 \pmod{p}, \\ &\qquad s = 2, 3, \ldots, p-2; \\ &\Sigma[1:h]D'_{p-2} + \frac{\Sigma[1:h]I(\alpha)}{2} \equiv 0 \pmod{p}, \end{aligned}\right\} \quad (7)$$

where

$$D'_s = \sum_{d=1}^{p-1} d^s(\alpha^d - \beta^b)$$

b some integer, $0 \leqq b < n-1$. A few transformations of (5) and (7) yield the

THEOREM I. *If* $x^p + y^p + z^p = 0$ *is satisfied in integers none zero and each prime to the odd prime* p, *then*

$$q(n)D_o \equiv 0, \qquad q(n)B_{\frac{s+1}{2}}D_s \equiv 0 \pmod{p},$$

$$s = 1, 3, \ldots, p-4,$$

where

$$D_s = \sum_{d=1}^{p-1} d^s \mathrm{I}(\alpha^d - \beta^a), \quad \left\{\frac{\theta}{\mathfrak{p}}\right\} = \alpha^{\mathrm{I}(\theta)},$$

$\mathfrak{p}$ *is a prime ideal divisor in the field* $\Omega(\alpha\beta)$ *of a rational odd prime* n *which is* $\not\equiv 0$ *or* $1 \pmod{p}$; $\alpha = e^{2i\pi/p}$; $\beta = e^{2i\pi/p}$; a *is some integer in the set* $1, 2, \ldots, n-2$ *other than* $(n-1)/2$; *the* B'*s are the numbers of Bernoulli,* $B_1 = 1/6$, $B_2 = 1/30$, etc.

Note that these criteria are independent of x, y and z.

THEOREM II. *If* $x^p + y^p + z^p = 0$ *is satisfied in integers none zero and each prime to the odd prime* p, *then*

$$q(n) \prod_{a=0}^{n-2} ((1-v)D_{p-2} + \mathrm{I}(\alpha)) \equiv 0 \pmod{p},$$

where v *has any one of the six values* $t, \frac{l}{t}, l-t, \frac{l}{l-t}, \frac{t-l}{t}, \frac{t}{t-l}$, $-x/y = t$ *the other symbols being defined as in Theorem I.*

THEOREM III. *If* p *is an odd prime and* $x^p + y^p + z^p = 0$ where $y \equiv 0 \pmod{p}$, $xz \not\equiv 0 \pmod{p}$, $x + z \not\equiv 0 \pmod{n}$, *then*

$$q(n)D_o \equiv 0$$

$$q(n)B_{\frac{s+1}{2}}D_s \equiv 0 \pmod{p},$$

$s = 1, 3, \ldots, p-4$; *and in addition one of the two relations*

$$q(n)D_{p-2} \equiv 0, \qquad q(n)\left(D_{p-2} + \frac{\mathrm{I}(\alpha)}{2}\right) \equiv 0 \pmod{p},$$

is satisfied, the other symbols being defined as in Theorem I.

Theorem II yields the

COROLLARY I. *If* $x^p + y^p + z^p = 0$ *is satisfied in integers none zero and all prime to the odd prime* p, *then*

$$q(n) \prod_{a=1}^{n-2} ((1-v)\ \mathrm{ind}\ (\alpha\beta^a - 1) - q(n)) \equiv 0 \pmod{p},$$

where

$$\left\{\frac{\alpha\beta^a - 1}{\mathfrak{q}}\right\} = \alpha^i, \quad i = \mathrm{ind}(\alpha\beta^a - 1),$$

$q = (\beta - r, n)$; *r a primitive root of n; v being defined as in Theorem II.*

COROLLARY II. *If* $x^p + y^p + z^p = 0$ *is satisfied in integers none zero and all prime to the odd prime p, then*

$$q(n) \prod_{a,b} (\text{ind}(\alpha\beta^a - 1)\text{ind}(\alpha\beta^b - 1) - q(n) (\text{ind}(\alpha\beta^a - 1) + \text{ind}(\alpha\beta^b - 1))) \equiv 0 \pmod{p},$$

where a and b each range independently over the integers, 1, 2, ..., $n - 2$, *the other symbols being defined as in Corollary I.*

Using the methods described above together with Furtwängler's law of reciprocity[3] we have the results:

THEOREM IV. *If* $xyz \not\equiv 0 \pmod{n}$, $z \not\equiv 0 \pmod{p}$, *n a prime* $\not\equiv 0 \pmod{p}$, *then* $q(n) \equiv 0 \pmod{p}$, or

$$\left.\begin{aligned} \sum_{s=2}^{p-2} (-1)^s k^s \left[\frac{d^s \log (x + e^v y)}{dv^s}\right]_{v=0} \left[\frac{d^{p-s} \log \gamma(e^v)}{dv^{p-s}}\right]_{v=0} \\ \equiv -\frac{kyh}{x + y} \mathrm{I}(\alpha) + h\mathrm{I}\left(\frac{\alpha^k - \beta^a}{1 - \beta^a}\right) \pmod{p}, \end{aligned}\right\} \quad (8)$$

where h is the integer to which the ideal $\mathfrak{Q}$ *belongs,* $\mathfrak{Q}$ *being a prime ideal divisor of n in the field* $\Omega(\alpha)$, $\mathfrak{Q}^h = (\gamma(\alpha))$, $\gamma(\alpha)$ *being an integer in* $\Omega(\alpha)$, *k is any integer* $\not\equiv 0 \pmod{p}$, *a is some integer in the set* 1, 2, ..., $n - 2$, *the other symbols being defined as in Theorem I, as criteria for the solution of* (1) *in integers, none zero.*

There is an analogous theorem for the case where $z \equiv 0 \pmod{p}$. The left-hand member of (8) obviously vanishes if x or $y \equiv 0 \pmod{p}$.

The methods indicated above yield many other results not stated here as well as at least four derivations of the Kummer criteria,[4] three proofs of Furtwängler's theorem,[5] the theorem of Sophie Germain[6] and in addition all the results contained in several papers by the writer.[7] THAT IS, THESE METHODS GIVE ALL RESULTS WHICH HAVE SO FAR LED TO THE EXCLUSION OF SPECIAL VALUES OF p IN

$$x^p + y^p + z^p = 0,$$

xyz NOT ZERO AND PRIME TO THE ODD PRIME p.

[1] Presented to the American Mathematical Society, Washington, D. C., Dec., 1924.

[2] *Ann. Math.*, **21**, 1919, p. 78.

[3] *Math. Ann.*, **72**, 1909, 386.

[4] Mirimanoff, *J. Math.*, **128**, 1905, 45–68.

[5] *Wiener Bericht*, Abt. IIa, **121**, 1912, 589–592.

[6] *Dickson's History of the Theory of Numbers*, Vol. II, p. 734.

[7] *Ann. Math.*, 2d ser., **26**, 1924–25, 88–94; 217–232; *Proc. Nat. Acad. Sci.*, **11**, 1925, 292–298.

MASS IN CURVED SPACE-TIME

BY G. Y. RAINICH

DEPARTMENT OF MATHEMATICS, JOHNS HOPKINS UNIVERSITY

Communicated December 16, 1925

The curvature of space-time has been introduced originally in order to explain gravitational phenomena and the approximate treatment of some special cases seems to have led to highly important results. But the question of the connection between the curvature tensor and the gravitation producing mass of a particle has not been treated, so far as I know, in an adequate way.

On the one hand, the four-dimensional formulation of hydrodynamic equations seems to suggest that mass, or, rather, density is somehow connected with the contracted Riemann tensor; but this suggestion does not seem to lead anywhere. On the other hand, in some special cases mass appears as a constant in the expression for the linear element (constant b in formula (1) below); the fact that this constant is equal or proportional to mass is established as a result of identifying the classical Newtonian equations with the first approximations of certain equations obtained from (1) in the case $a = 0$, but it has not been shown how this constant b can be derived from the curvature.

In the domain of electromagnetism which seemed at first not to be included in the scheme of the general relativity theory we know now the answer to the corresponding question of the connection between the curvature tensor and the electric charge.[1] The latter appears as the residue of a field derived from the curvature field, and it is natural to expect to have a similar expression for the mass. Nordström[2] obtained such an expression but his integrand is a *quasi*-vector, and a new treatment of the question seems to be desirable.

In the present note an expression of the type of a residue is given for mass in the centrosymmetric case. The questions of how the same method can be applied in other cases and of what are the restrictions imposed on space-time by the existence of mass will be treated later but already now it seems to be possible to formulate two conclusions of a general nature which follow from the special considerations given below.

One conclusion is that the mass constant appears to be connected *not* with the *contracted* but with the *complete* Riemann tensor.

The other is that in order to introduce mass we have to *separate space and time.*

1. We first summarize the situation as to the electric charge. Given the curvature tensor we form the contracted curvature tensor, we obtain the electromagnetic tensor by performing certain algebraic operations on

the contracted tensor and solving a differential equation, and then the square of the electric charge is obtained as the value of certain integral expressions taken over a surface surrounding the singularity, the integrands being given by the electromagnetic tensor and the value of the expressions being independent of the choice of the surface provided it surrounds the same singularity. In the special case when the line element is given by

$$-ds^2 = \xi(r)\,dr^2 + r^2 d\vartheta^2 + r^2 \sin^2\vartheta d\psi^2 - \eta(r)dt^2 \tag{1}$$

with

$$\eta(r) = 1/\xi(r) = 1 + b/r + a/r^2 \tag{2}$$

the complete (not contracted) curvature tensor can be given by its components

$$\begin{gathered} R_{14}^{14} = -b/r^3 - 3a/r^4, \quad R_{23}^{23} = -b/r^3 - a/r^4, \\ R_{12}^{12} = R_{13}^{13} = R_{24}^{24} = R_{34}^{34} = b/2r^3 + a/r^4; \end{gathered} \tag{3}$$

the mixed components of the contracted tensor are

$$R_1^1 = R_4^4 = -a/r^4, \quad R_2^2 = R_3^3 = a/r^4, \tag{4}$$

all other components being zero in both cases. We see that the components of the contracted tensor involve a only, that b has been eliminated by the process of contraction. In this special case the general process for the calculation of the charge outlined above reduces simply to the integration of the vector with the components $\pm\sqrt{a}/r^2,0,0,0$ over any surface surrounding the line $r = 0$, e.g., over a sphere $r = \text{const.}$, $t = \text{const.}$ The constant a is thus identified as the square of the charge.[3]

2. As mentioned above the constant b is connected with the *mass* of the particle represented by the singularity $r = 0$ (in appropriate units we can write $b = -2m$); it seems to be natural to try to separate out this constant b by the application of some process similar to contraction which helped us to separate out a and eliminate b when we were interested in the electric charge.

It appears, however, to be impossible to obtain from the curvature tensor R_{kl}^{ij} a tensor involving (explicitly) b only, by the application of an *intrinsic* operation, i.e., an operation which does not give preference to any direction. We shall try, therefore, to find an operation which eliminates a and which is *not* intrinsic; as a preferred direction we choose the time direction t, or, what amounts to the same thing, we consider at each point the space components only of the Riemann tensor and the contracted Riemann tensor. They are

$$R_{23}^{23} = -b/r^3 - a/r^4, \quad R_{31}^{31} = R_{12}^{12} = b/2r^3 + a/r^4; \tag{5}$$

$$R_1^1 = -a/r^4, \quad R_2^2 = R_3^3 = a/r^4. \tag{6}$$

Subtracting we find a tensor (of the three-space) with the components

$$m_1^1 = b/r^3, \qquad m_2^2 = m_3^3 = -b/2r^3 \tag{7}$$

which do not contain a.

3. We have attained thus our aim but at the sacrifice of intrinsicity. If we wish now to consider the tensor *field* m_j^i we naturally have to consider it on one of the spaces t = constant—the separation of time and space directions at a point leads to the separation of time and space at large.

In flat space-time a tensor field with the components (7) is the differential of a vector field with the components

$$f^1 = -b/2r^2, \qquad f^2 = f^3 = 0. \tag{8}$$

We recognize the *Newtonian field of force.*

In our curved space the contravariant components of the corresponding vector (i.e., a vector of the same length $-b/2r^2$ in the direction r) are

$$F^1 = -\sqrt{\eta}\, b/2r^2, \qquad F^2 = F^3 = 0. \tag{9}$$

The differential of this vector field has for its mixed components

$$M_1^1 = \sqrt{\eta}\, b/r^3, \qquad M_2^2 = M_3^3 = -\sqrt{\eta}\, b/2r^3, \tag{10}$$

all the others zero.

If we integrate our field of force over a surface surrounding the singular point $r = 0$ we obtain $4\pi m$ so that the mass constant is, as was to be expected, the residue of a field derived from the curvature field.

The process then by which we obtain the mass constant from the curvature field in the centro-symmetric case consists in the following steps: *we form the difference between the space components of the contracted and the complete Riemann tensors; we multiply the result by* $\partial s/\partial t$; *we consider the product as the differential of a vector field in space; the residue of this vector field is the mass constant.*

It must be noted that there is no ambiguity in the sign of the residue; the square roots are cancelled out in the result; that is why we have only one kind of masses, whereas we have positive and negative electricity because the extraction of a square root is essentially involved in deducing the electromagnetic tensor from the Riemann tensor, or in calculating the electric charge.

4. In the "special relativity theory" mass is considered as changing with velocity or as depending on the choice of separation of time and space.

There does not seem to exist an obvious modification of the whole process described in sect. 3 which would permit us to apply it for a separation of time and space different from the one corresponding to t = const. used above. But it may be that there are other separations for which $\partial s/\partial t'$ (t' denoting the new time) differs from $\partial s/\partial t$ by a constant factor. If

we use in the second step this $\partial s/\partial t'$ instead of the original $\partial s/\partial t$ the mass will be multiplied by the constant factor and this may correspond to the change of mass considered in the special relativity theory.

[1] Cf. a paper by the present writer, *New York, Trans. Amer. Math. Soc.*, **27**, 1925, January, pp. 106–136.

[2] *Amsterdam, Proceedings*, **20**$_2$, 1918, pp. 1076–1091.

[3] Cf. paper cited under[1] sections 9 and 12.

CONCERNING CERTAIN SOLVABLE EQUATIONS WITH FUNCTIONAL DERIVATIVES

BY ARISTOTLE D. MICHAL*

CAMBRIDGE, MASS.

Communicated December 9, 1925

In this note I shall consider briefly the functional equations arising in connection with a generalization of the interesting Appell polynomials.[1] In a paper to be published elsewhere this work with additional material will appear in certain detail.

Let $F^{(i)}$ stand for the ith degree polynomial functional

$$f_0^{(i)} + \sum_{k=1}^{i} \int_0^1 \int_{0(k)}^1 \cdots \int_0^1 f_k^{(i)}(t_1,t_2,\ldots,t_2)y(t_1)y(t_2)\ldots y(t_k)dt_1dt_2\ldots dt_k. \quad (1)$$

The argument $y(x)$ of the functional $F^{(i)}$ is assumed continuous in the interval $(0, 1)$. We also assume that the functions $f_i^{(k)}(t_1\ldots,t_k)$ are continuous in the intervals $0 \leq t_j \leq 1$, and, without loss of generality, symmetric in all their arguments.

Denoting by $F_y^{(i)}(t)$ the first functional derivative of $F^{(i)}$ taken at the point t, we state the following fundamental theorem.

THEOREM I. *A necessary and sufficient condition that a set of polynomial functionals* $F^{(i)}$ $(i = 0, 1, 2,\ldots,n)$ *has the property*

$$F_y^{(i)}(t) = i\varphi(t)F^{(i-1)} \qquad (i = 1, 2,\ldots,n) \quad (2)$$

($\varphi(t)$ *being a given arbitrary continuous function in the interval* $(0, 1)$) is that

$$\begin{cases} F^{(i)} = g_{i-1} + \sum_{k=1}^{i} \binom{i}{k} g_{i-k}\left[\int_0^1 \varphi(t)y(t)dt\right]^k \\ F^{(0)} = g_0 \end{cases} \quad (i = 1, 2,\ldots,n). \quad (3)$$

Consider the formal expression $g(h)$ defined by

$$g(h) = \sum_{k=0}^{\infty} \frac{g_k h^k}{k!}. \quad (4)$$

It is readily verified that the coefficient of $h^i/i!$ in the formal expression

$$g(h)e^{h\int_0^1 \varphi(t)y(t)dt} \tag{5}$$

is precisely the ith degree polynomial functional[2] $F^{(i)}$. We shall call $g(h)$ the generating function and $\varphi(t)$ the associated function of the set of polynomial functionals $F^{(i)}$.

Let

$$F_1^{(0)}, F_1^{(1)}, \ldots, F_1^{(n)}, \ldots \qquad F_2^{(0)}, F_2^{(1)}, \ldots, F_2^{(n)}, \ldots \tag{6}$$

be two sets of polynomial functionals having $g_1(h)$ and $g_2(h)$ as generating functions, and $\varphi_1(t)$ and $\varphi_2(t)$ as associated functions, respectively. Denote by the notation

$$(F_1F_2)^{(0)}, \quad (F_1F_2)^{(1)}, \ldots, (F_1F_2)^{(n)}, \ldots \tag{7}$$

the set of polynomial functionals formed by putting $F_2^{(i)}$ for $\left[\int_0^1 \varphi_1(t)y(t)dt\right]$ in the set $F_1^{(0)}, \ldots, F_1^{(n)}, \ldots$ This leads us to the theorem

THEOREM II. *The set of polynomial functionals* (7) *will have* $g_1(h)\,g_2(h)$ *as its generating function and* $\varphi_2(t)$ *as its associated function.*

If we have two sets of polynomial functionals (6) with the same associated function $\varphi(t)$ such that

$$(F_1F_2)^{(i)} = \left[\int_0^1 \varphi(t)y(t)dt\right]^i \tag{8}$$

then we define the polynomial functionals F_2 as the inverses of the polynomial functionals F_1. We shall denote these inverses by the symbol (F_1^{-1}). The generating function for the inverses (F_1^{-1}) is the reciprocal of the generating function for the F_1's.

Let $(d^kF)^{(i)}$ represent a polynomial functional of the ith degree which belongs to a set of polynomial functionals having the formal expression $d^kg(h)/dh^k$ as a generating function and $\varphi(t)$ as the associated function. We shall need to refer to the formula

$$(d^kF)^{(i)} = F^{(i+k)} - \binom{k}{1}\left[\int_0^1 \varphi(t)y(t)dt\right]F^{(i+k-1)} + \binom{k}{2}\left[\int_0^1 \varphi(t)y(t)dt\right]^2 F^{(i+k-2)} - \ldots, \tag{9}$$

where in the second member of this equality, all terms in $\int_0^1 \varphi(t)y(t)dt$ of degree higher than i cancel so that $(d^kF)^{(i)}$ is a polynomial functional of degree i.

We proceed now in outlining a method of deducing the functional equations with functional derivatives satisfied by polynomial functionals in

the cases in which the generating function $g(h)$ of the set of polynomial functionals F satisfies formally a linear differential equation with coefficients which are polynomials in h. Multiplying this differential equation by

$$e^{h\int_0^1 \varphi(t)y(t)dt}$$

and replacing the expressions

$$\frac{d^i g(h)}{dh^i} \cdot e^{h\int_0^1 \varphi(t)y(t)dt}$$

by the expressions of definition

$$\sum_{j=0} \frac{h^j}{j!} (d^i F)^{(j)},$$

respectively, we shall get, on equating coefficients of h^i, a linear relation between the F's, (dF)'s,$\ldots$,(d^kF)'s,$\ldots$. Finally employing relations (9) and (2) we shall get the desired functional equation. We give the following illustration. Let the hypergeometric series of Gauss $H(\alpha,\beta,\gamma,h)$ be the generating function and $\varphi(t)$ the associated function. The generating function H satisfies the differential equation

$$(h - h^2)\frac{d^2H}{dh^2} + (\gamma - \alpha h)\frac{dH}{dh} - bH = 0.$$

Putting $\alpha + \beta + 1 = a$, $\alpha\beta = b$, we deduce, on following out the above method, the functional equation with third order functional derivatives

$$\varphi(t_1)\varphi(t_2)\varphi(t_3)(i+\gamma-1)F^{(i)} - \varphi(t_2)\varphi(t_3)\Big[(i-1)(2\int_0^1 \varphi(t)y(t)dt + a + i - 2) + \gamma\int_0^1 \varphi(t)y(t)dt + b\Big]F_y^{(i)}(t_1) + \varphi(t_3)\int_0^1 \varphi(t)y(t)dt\Big(\int_0^1 \varphi(t)y(t)dt + a + 2i - 4\Big)F_{yy}^{(i)}(t_1, t_2) - \Big[\int_0^1 \varphi(t)y(t)dt\Big]^2 F_{yyy}^{(i)}(t_1, t_2, t_3) = 0$$

satisfied by the polynomial functional of the ith degree.

We note in passing that we can consider analogous problems in the case of functionals of more than one independent function. For example, in the case of functionals of two independent functions $y(x)$ and $z(x)$ we can consider relations analogous to (2) such as

$$F_y^{(m,n)}(t) = am\varphi_1(t)F^{(m-1,n)}, + bn\varphi_2(t)F^{(m,n-1)}$$
$$F_z^{(m,n)}(t) = a^1m\varphi_3(t)F^{(m-1,n)} + b^1n\varphi_4(t)F^{(m,n-1)}.$$

We turn now briefly to the analogous theories of functionals[3] $F_{[c]}^{(i)}$ of closed curves C of type

$$\varphi_0^{(i)} + \sum_{k=1}^{i} \int_\sigma \int_\sigma \cdots \int_\sigma \varphi_k^{(i)}(M_1, M_2, \ldots, M_k) d\sigma_1, d\sigma_2 \ldots d\sigma_k. \quad (10)$$

The following theorem is the key to the situation.

THEOREM III. *A necessary and sufficient condition that a set of $n + 1$ functionals of closed curves of type (10) has the property given by the relations*

$$F_n^{(i)}(M) = i\psi(M) F_{[c]}^{(i-1)}$$

($F_n^{(i)}(M)$ standing for the normal functional derivative of $F_{[c]}^{(i)}$) is that

$$\begin{cases} F_{[c]}^{(i)} = g_{i-1} + \sum_{k=1}^{i} \binom{i}{k} g_{i-k} \left[\int_\sigma \psi(M) d\sigma\right]^k & (i = 1, 2, \ldots, n) \\ F_{[c]}^{(0)} = g_0. \end{cases}$$

Finally we consider the corresponding generalizations in Evans' functional algebra.[4] Let

$$\xi = u + jU(r, s),$$

as in Evans' algebra. We ask the question whether it is possible to define differentiation in Evans' algebra[5] so that the first such operation applied to the expression $F_{(\xi)}^{(i)}$ of the form

$$a_0 + a_1\xi + a_2\xi^2 + \ldots + a_i\xi^i \quad (11)$$

yields

$$a_1 + 2a_2\xi + \ldots + ia_i\xi^{i-1}. \quad (12)$$

An affirmative answer is implied by the following theorem.

THEOREM IV. *Let $\bar{\xi} = \bar{u} + j\,\bar{U}(r, s)$, $\bar{U}(r, s)$ being any finite continuous function permutable with the functions $U(r, s)$. The limit* $\lim_{\epsilon=0} \dfrac{F_{(\xi+\epsilon\bar{\xi})}^{(i)} - F_{(\xi)}^{(i)}}{\epsilon\bar{\xi}}$ *exists, is independent of ξ and is equal to the expression (12).*

The proof of this theorem depends on the properties of reciprocal functions in integral equation theory.

Thus we see that the Appell polynomial theory has an isomorph in Evans' functional algebra when differentiation in Evans' algebra is defined as stated in Theorem IV.

* NATIONAL RESEARCH FELLOW IN MATHEMATICS.

[1] See Appell, *Ann. Ecole Normal* (1880), pp. 119–144.

[2] By a polynomial functional we shall mean, from now on, a functional of type (3).

[3] The extension to functionals of closed $n-1$ dimensional spreads in n dimensions is clear.

[4] Cf. G. C. Evans, *Cambridge Colloquium Lectures on Mathematics*, pp. 119–121.

[5] To get a unique result in division (the solution of an integral equation being involved) we assume that we are considering the algebra with the variable limits. We can, however, treat the algebra with constant limits if we restrict ourselves to a certain type of functions $U(r, s)$. Cf. Evans, loc. cit.

FIRST INTEGRALS IN THE GEOMETRY OF PATHS

BY JOSEPH MILLER THOMAS*

PRINCETON UNIVERSITY

Communicated December 30, 1925

1. The object of this paper is to discuss the conditions which must be fulfilled in order that the differential equations whose properties constitute the geometry of paths[1] have first integrals of particular types. In §2 three general theorems concerning first integrals of the rth degree are proved. In §3 a geometrical necessary and sufficient condition for the existence of a linear first integral is given.

The subsequent sections are devoted to quadratic first integrals. Such an integral gives a tensor of the second rank, which can be adopted as the metric tensor of the space. In §4 a necessary and sufficient condition that an affine geometry of paths be a Weyl metric geometry is derived.

In §5 I obtain a necessary and sufficient condition that one of the affine geometries of a given system of paths be a Weyl metric geometry. In order to apply the theorem, however, it is in general necessary to determine whether the affine connection can be chosen so that the corresponding equations of the paths have a quadratic first integral. As this projective problem of the quadratic first integral is as yet unsolved, the condition cannot be regarded as entirely satisfactory. It is, nevertheless, useful in certain cases; in particular, when we wish to determine whether a geometry of paths admitting a quadratic first integral is projectively equivalent to a Weyl metric geometry.

Finally (§6) I derive a necessary and sufficient condition that one of the affine geometries of a given system of paths be Riemannian, and consequently solve the projective problem corresponding to an affine problem which has already been solved.[2] The conditions here found presuppose a method of determining whether a given form of the equations of the paths admits a quadratic first integral. Such a criterion has already been established.[3]

2. It is known[4] that the following simultaneous transformations of the affine connection and the parameter leave the paths unaltered:

$$\Gamma^i_{jk} = \Gamma'^i_{jk} + \delta^i_j\varphi_k + \delta^i_k\varphi_j, \tag{2.1}$$

$$\exp.\ (-2\int \varphi_i dx^i)\frac{ds}{dt} = 1. \tag{2.2}$$

Here, as throughout this paper, the integral is understood to be evaluated along the path in question.

If the original form of the equations of the paths, namely,

$$\frac{d^2x^i}{ds^2} + \Gamma^i_{jk}\frac{dx^j}{ds}\frac{dx^k}{ds} = 0, \tag{2.3}$$

admits the following homogeneous first integral of the rth degree

$$g_{ij\ldots k}\frac{dx^i}{ds}\frac{dx^j}{ds}\cdots\frac{dx^k}{ds} = \text{const.}, \tag{2.4}$$

then by multiplying (2.4) by (2.2) r times we see that the new form of the equations

$$\frac{d^2x^i}{dt^2} + \Gamma'^i_{jk}\frac{dx^j}{dt}\frac{dx^k}{dt} = 0 \tag{2.5}$$

admits the first integral

$$\exp.\,(-2r\int\varphi_p dx^p)g_{ij\ldots k}\frac{dx^i}{dt}\frac{dx^j}{dt}\cdots\frac{dx^k}{dt} = \text{const.} \tag{2.6}$$

Conversely, if equations (2.3) admit a first integral of the form

$$\exp.\,(\int\psi_p dx^p)g_{ij\ldots k}\frac{dx^i}{ds}\frac{dx^j}{ds}\cdots\frac{dx^k}{ds} = \text{const.},$$

where ψ_p denotes a vector, by taking φ_p in (2.1) equal to $\psi_p/2r$ we obtain an affine connection for which equations (2.5) admit the first integral

$$g_{ij\ldots k}\frac{dx^i}{dt}\frac{dx^j}{dt}\cdots\frac{dx^k}{dt} = \text{const.} \tag{2.7}$$

It can be assumed that the tensor $g_{ij\ldots k}$ is symmetric in all the indices. Let the covariant derivative of such a tensor with respect to the Γ's be denoted by $g_{ij\ldots k,l}$ and its covariant derivative with respect to the Γ''s by $(g_{ij\ldots k,l})'$. We then have

$$\begin{aligned}(g_{ij\ldots k,l})' &= g_{ij\ldots k,l} + (r-1)g_{ij\ldots k}\varphi_l + P(g_{ij\ldots k}\varphi_l),\\ P[(g_{ij\ldots k,l})'] &= P(g_{ij\ldots k,l} + 2r\,g_{ij\ldots k}\varphi_l),\end{aligned} \tag{2.8}$$

where the symbol P denotes the sum of the terms obtained from those in the parentheses by cyclic permutation of the indices.

Veblen and T. Y. Thomas[5] give the following relations

$$P[(g_{ij\ldots k,l})'] = 0$$

as necessary and sufficient conditions that (2.7) be a first integral of (2.5). By means of (2.8) these conditions can be written

$$P(g_{ij\ldots k,l} + 2r\,g_{ij\ldots k}\varphi_l) = 0. \tag{2.9}$$

Hence we have the following theorem originally given, in substance, for $r = 2$ by Eisenhart.[6]

THEOREM 1. *A necessary and sufficient condition that the affine connection can be chosen so that the corresponding equations of the paths admit a homogeneous first integral of the rth degree is the existence of a symmetric tensor* $g_{ij\ldots k}$ *and a vector* ψ_i *satisfying the relations*

$$P(g_{ij\ldots k,l} + g_{ij\ldots k}\psi_l) = 0.$$

If these conditions are fulfilled, the first integral is

$$g_{ij\ldots k}\frac{dx^i}{dt}\frac{dx^j}{dt}\cdots\frac{dx^k}{dt} = \text{const.},$$

and the corresponding affine connection is obtained from the original by the formulas

$$\Gamma'^i_{jk} = \Gamma^i_{jk} - \delta^i_j\varphi_k - \delta^i_k\varphi_j, \qquad 2r\varphi_j = \psi_j.$$

In order that the coefficients in (2.6) reduce to point functions, it is necessary and sufficient that

$$\int \varphi_p\, dx^p = \sigma(x,x_o),$$

where x_o denotes the initial point. Differentiating along the curve, x_o remaining constant, we get

$$(\varphi_p - \partial\sigma/\partial x^p)dx^p = 0.$$

Considering n paths whose tangent vectors are linearly independent at the point x, we get $\varphi_p = \partial\sigma/\partial x^p$; whence by differentiation with respect to x^q_o, we find $\partial^2\sigma/\partial x^p\partial x^q_o = 0$, or $\sigma = \varphi(x) + F(x_o)$. Hence φ_i is the gradient of a scalar φ. The factor $\exp[-2rF(x_o)]$ can be absorbed in the constant of the right-hand member of (2.6).

THEOREM 2. *A projective change of the affine connection carries a homogeneous first integral of the equations of the paths into another such integral if, and only if, the vector defining the change of the affine connection is the gradient of a scalar.*

The following theorem, a consequence of theorems 1 and 2, embodies and extends a theorem given by Veblen and T. Y. Thomas.[7]

THEOREM 3. *A necessary and sufficient condition that the equations of the paths corresponding to a given affine connection admit a homogeneous first integral of the rth degree is the existence of a symmetric tensor* $g_{ij\ldots k}$ *and a scalar* ψ *which satisfy the relations*

$$P(g_{ij\ldots k,l} + g_{ij\ldots k}\psi_l) = 0, \qquad \psi_l = \partial\psi/\partial x^l.$$

If these conditions are satisfied, the first integral is

$$e^{\psi}g_{ij\ldots k}\frac{dx^i}{ds}\frac{dx^j}{ds}\cdots\frac{dx^k}{ds} = \text{const.},$$

whereas the tensor $g_{ij\ldots k}$ itself is the coefficient of a homogeneous first integral of the equations of the paths for the affine connection Γ'^{i}_{jk} defined by the formulas

$$\Gamma'^{i}_{kj} = \Gamma^{i}_{jk} - \delta^{i}_{j}\varphi_k - \delta^{i}_{k}\varphi_j, \qquad 2r\varphi_j = \psi_j.$$

3. Suppose that the vector a_i gives a linear first integral for the connection Γ'. From (2.9) we have

$$a_{i,j} + a_{j,i} + 2a_i\varphi_j + 2a_j\varphi_i = 0.$$

By choosing $2\varphi_i = a_i$ we get an affine connection Γ satisfying the relations

$$a_{i,j} + a_{j,i} + 2a_ia_j = 0. \tag{3.1}$$

Equations (3.1) are the conditions that the normal coördinates for Γ and Γ'', where $\Gamma''^{i}_{jk} = \Gamma^{i}_{jk} - \delta^{i}_{j}a_k - \delta^{i}_{k}a_j$, be related by a linear fractional transformation at every point.[8] Conversely, if equations (3.1) are satisfied, a_i furnishes a linear first integral for the equations of the paths corresponding to the affine connection $\Gamma'^{i}_{jk} = \Gamma^{i}_{jk} - {}^1/_2\delta^{i}_{j}a_k - {}^1/_2\delta^{i}_{k}a_j$.

THEOREM 4. *In order that some form of the equations of the paths admit a linear first integral it is necessary and sufficient that the normal coördinates for two distinct affine connections be related by a linear fractional transformation at every point.*

4. Let us next consider when an affine connection Γ admits a Weyl metric with fundamental tensor g_{ij} and vector ψ_i. The conditions for this are[9]

$$g_{ij,k} + g_{ij}\psi_k = 0. \tag{4.1}$$

The integrability conditions of this system of equations can be treated by slight modifications of a method already developed.[10] By covariant differentiation of equations (4.1) and use of the well known identity

$$g_{ij,k,l} - g_{ij,l,k} = -g_{\alpha j}B^{\alpha}_{ikl} - g_{i\alpha}B^{\alpha}_{jkl}, \tag{4.2}$$

B^{i}_{jkl} denoting the curvature tensor, we easily eliminate ψ_i and their derivatives. Differentiating the resulting equations successively and combining with (4.1) we find that any tensor g_{ij} which satisfies (4.1) must also satisfy the infinite sequence of equations (4.7) which are given below in theorem 5.

Suppose that the first N systems of the equations (4.7) admit a fundamental set of solutions g^{μ}_{ij} which satisfy the $(N+1)$th system. The general solution of the first N systems will be

$$g_{ij} = \varphi^{\mu}g^{\mu}_{ij}, \tag{4.3}$$

where φ^{μ} are r scalars. It is to be noted that μ is not an index which denotes tensor character, but is one which distinguishes the different fundamental solutions.

It is easily shown that the algebraic solutions of (4.7) have tensor character. From the assumptions concerning g^{μ}_{ij} it is then readily proved that

$$g^{\mu}_{ij,k} = \lambda^{\mu\sigma}_{k}\, g^{\sigma}_{ij}, \tag{4.4}$$

where the λ's are r^2 covariant vectors.

In order to get a solution of our problem, we must determine the scalars φ^{μ} so that (4.3) satisfy (4.1). The differential equations which φ^{μ} must satisfy are found by direct substitution to be

$$\partial\varphi^{\mu}/\partial x^{k} + \varphi^{\sigma}\lambda^{\sigma\mu}_{k} + \varphi^{\mu}\psi_{k} = 0.$$

The integrability conditions of these equations are

$$\varphi^{\sigma}(\partial\lambda^{\sigma\mu}_{k}/\partial x^{l} - \partial\lambda^{\sigma\mu}_{l}/\partial x^{k} + \lambda^{\sigma\tau}_{k}\lambda^{\tau\mu}_{l} - \lambda^{\sigma\tau}_{l}\lambda^{\tau\mu}_{k}) + \varphi^{\mu}(\partial\psi_{k}/\partial x^{l} - \partial\psi_{l}/\partial x^{k}) = 0. \tag{4.5}$$

Since g^{σ}_{ij} satisfies the first system of (4.7) and also (4.2) we have

$$g^{\sigma}_{ij,k,l} - g^{\sigma}_{ij,l,k} = -2g^{\sigma}_{ij}\, S_{kl}/n,$$

and from (4.4) we get

$$g^{\sigma}_{ij,k,l} = \lambda^{\sigma\mu}_{k,l}g^{\mu}_{ij} + \lambda^{\sigma\tau}_{k}\lambda^{\tau\mu}_{l}g^{\mu}_{ij}.$$

Combination of the last two sets of equations gives

$$g^{\mu}_{ij}(\partial\lambda^{\sigma\mu}_{k}/\partial x^{l} - \partial\lambda^{\sigma\mu}_{l}/\partial x^{k} + \lambda^{\sigma\tau}_{k}\lambda^{\tau\mu}_{l} - \lambda^{\sigma\tau}_{l}\lambda^{\tau\mu}_{k} + 2\delta^{\sigma}_{\mu}S_{kl}/n) = 0.$$

Since the quantities g^{μ}_{ij} are linearly independent, the coefficients in these last equations must be identically zero. Combining this result with (4.5), we find that the conditions of integrability of the equations for φ^{μ} are

$$\partial\psi_{k}/\partial x^{l} - \partial\psi_{l}/\partial x^{k} = 2S_{kl}/n. \tag{4.6}$$

Eisenhart[11] has shown that S_{kl} is always the curl of a vector. Hence ψ_i can always be chosen so that (4.6) are satisfied.

THEOREM 5. *A necessary and sufficient condition that an affine geometry of paths be a Weyl metric geometry is the existence of a positive integer N such that the first N systems of the equations*

$$\begin{aligned} g_{\alpha\beta}G^{\alpha\beta}_{ijkl} &= 0, \\ g_{\alpha\beta}G^{\alpha\beta}_{ijkl,p} &= 0, \\ g_{\alpha\beta}G^{\alpha\beta}_{ijkl,p,q} &= 0, \\ \cdots \quad \cdots \quad \cdots \quad \cdots, \end{aligned} \tag{4.7}$$

where

$$G^{\alpha\beta}_{ijkl} = \delta^{\alpha}_{i}(B^{\beta}_{jkl} - \delta^{\beta}_{j}S_{kl}/n) + \delta^{\alpha}_{j}(B^{\beta}_{ikl} - \delta^{\beta}_{i}S_{kl}/n),$$

and

$$S_{kl} = B^{\alpha}_{\alpha kl},$$

admit a fundamental system of solutions which satisfy the $(N + 1)$*th system of equations.*

5. The affine connection Γ' admits a Weyl metric provided that

$$(g_{ij,k})' + g_{ij}\psi_k = 0. \tag{5.1}$$

From theorem 1 it is seen that a necessary condition is that g_{ij} furnish a quadratic first integral for a properly chosen form of the equations of the paths. Let us suppose that Γ^i_{jk} is the affine connection corresponding to such a choice. Theorem 1 then gives the following relations

$$\psi_k = -4\varphi_k, \tag{5.2}$$

$$\Gamma'^i_{jk} = \Gamma^i_{jk} - \delta^i_j\varphi_k - \delta^i_k\varphi_j,$$

by the use of which (5.1) become

$$g_{ij,k} + g_{jk}\varphi_i + g_{ik}\varphi_j - 2g_{ij}\varphi_k = 0. \tag{5.3}$$

We easily obtain

$$(n-1)\varphi_k = {}^1/_{2g}{}^{\alpha\beta}g_{\alpha\beta,k} = \frac{\partial \log \sqrt{g}}{\partial x^k} - \Gamma^{\alpha}_{\alpha k}, \tag{5.4}$$

where $g = |g_{ij}|$. By substitution in (5.3) we get

$$g_{ijk} = 0, \tag{5.5}$$

where

$$g_{ijk} = 2(n-1)g_{ij,k} + g^{\alpha\beta}(g_{jk}g_{\alpha\beta,i} + g_{ik}g_{\alpha\beta,j} - 2g_{ij}g_{\alpha\beta,k}). \tag{5.6}$$

Conversely, if the affine connection Γ^i_{jk} admits a quadratic first integral g_{ij} which satisfies (5.5), the vector φ_k defined by (5.4) is such that g_{ij}, $\Gamma'^i_{jk} = \Gamma^i_{jk} - \delta^i_j\varphi_k - \delta^i_k\varphi_j$, and $\psi_k = -4\varphi_k$ satisfy (5.1).

THEOREM 6. *A necessary and sufficient condition that one of the affine geometries of a system of paths be a Weyl metric geometry is that the equations of the paths corresponding to some affine connection admit a quadratic first integral whose fundamental tensor* g_{ij} *satisfies the relations*

$$2(n-1)g_{ij,k} + g^{\alpha\beta}(g_{jk}g_{\alpha\beta,i} + g_{ik}g_{\alpha\beta,j} - 2g_{ij}g_{\alpha\beta,k}) = 0.$$

6. As is well known, the Weyl geometry reduces to a Riemann geometry if, and only if the vector ψ_i is the gradient of a scalar. From this fact and equations (5.2) and (5.4) it follows that the quadratic first integral corresponding to a Riemann geometry must belong to an affine connection for which the skew-symmetric tensor S_{ij} vanishes

$$S_{ij} \equiv \frac{\partial \Gamma^{\alpha}_{\alpha i}}{\partial x^j} - \frac{\partial \Gamma^{\alpha}_{\alpha j}}{\partial x^i} = 0. \tag{6.1}$$

Eisenhart[12] has shown that if we define a vector φ_i by the equations

$$\varphi_i = \frac{1}{2(n+1)} b^{\alpha\beta} b_{\alpha\beta,i} \tag{6.2}$$

where b_{ij} denotes the symmetric part of the Ricci tensor formed for the Γ's, the affine connection Γ' satisfies (6.1).

The equations of the paths corresponding to an affine connection which satisfies (6.1) will be said to be in *reduced form*. It is easily shown that any two affine connections satisfying (6.1) are related by equations of the type (2.1) in which φ_i is the gradient of a scalar. Hence theorem 2 shows that every reduced form of the equations has a quadratic first integral provided one such form does.

Applying theorem 6 to the present case, we have, therefore,

THEOREM 7. *A necessary and sufficient condition that one of the affine geometries of a system of paths be Riemannian is that every reduced form of the equations of the paths admit a quadratic first integral whose fundamental tensor g_{ij} satisfies the equations*

$$2(n-1)g_{ij,k} + g^{\alpha\beta}(g_{jk}g_{\alpha\beta,i} + g_{ik}g_{\alpha\beta,j} - 2g_{ij}g_{\alpha\beta,k}) = 0.$$

The process of testing whether a given geometry of paths admits a Riemann metric is the following. The first step is to put the equations of the paths in reduced form. From equations (6.2) we see that the carrying out of this involves only algebraic operations and differentiation. The method developed by Veblen and T. Y. Thomas[13] is then applied to the reduced form of the equations, and by algebraic operations it is determined whether these equations have a quadratic first integral or not. In case such an integral exists, the tensor g_{ij} and its covariant derivative $g_{ij,k}$ are found by integration of a completely integrable system of the first order. These values are then to be substituted in (5.5). If the arbitrary constants in the solution can be chosen so that (5.5) are satisfied, the Riemann metric exists; otherwise it is impossible to associate such a metric with the paths. In the case of their existence, the Christoffel symbols of the second kind are obtained from (2.1) and (5.4). The fundamental tensor is given by

$$g'_{ij} = e^{-4\varphi} g_{ij}, \tag{6.3}$$

where

$$\frac{\partial \varphi}{\partial x^i} = \varphi_i.$$

7. It is worthy of note that the tensor g_{ijk} defined by (5.6) is an invariant of the g's and the Γ's which under simultaneous transformation

of the tensor g_{ij} and the affine connection by means of (6.3) and (2.1) has the following law of transformation

$$g'_{ijk} = e^{-4\varphi} g_{ijk}.$$

8. Let us apply some of the preceding results to the problem of geodesic correspondence between two Riemann spaces whose fundamental tensors are g_{ij} and γ'_{ij}. In order that such a correspondence exists, it is necessary and sufficient that the Christoffel symbols of the second kind for the two spaces be related by equations (2.1) in which φ_i is necessarily the gradient of a scalar.

Theorem 2 shows that the equations of the geodesics of the first space have the first integral

$$\gamma_{ij} \frac{dx^j}{ds} \frac{dx^i}{ds} = \text{const.},$$

and those of the second, the integral

$$g'_{ij} \frac{dx^i}{dt} \frac{dx^j}{dt} = \text{const.},$$

where

$$\gamma'_{ij} = e^{-4\varphi} \gamma_{ij}, \qquad g'_{ij} = e^{-4\varphi} g_{ij}. \tag{8.1}$$

This result is well known.

Theorem 7 applied to the present case shows that γ_{ij} must satisfy equations of the type (5.3). If we substitute in these equations the value of γ_{ij} given by (8.1), we get the conditions on γ'_{ij} given by Levi-Civita.[14]

* NATIONAL RESEARCH FELLOW IN MATHEMATICS.

[1] Cf. O. Veblen and T. Y. Thomas, *Trans. Amer. Math. Soc.*, **25**, 1923, pp. 551–580.

[2] L. P. Eisenhart and O. Veblen, these PROCEEDINGS, **8**, 1922, p. 19, find a *sufficient* condition, and Veblen and T. Y. Thomas, loc. cit., p. 591, find a *necessary and sufficient* condition that a *given* affine connection Γ^i_{jk} be the Christoffel symbols of the second kind of a Riemannian space.

[3] Veblen and T. Y. Thomas, loc. cit., pp. 599–608.

[4] Veblen and T. Y. Thomas, loc. cit., p. 557. The result was first obtained by Weyl.

[5] Loc. cit., p. 582.

[6] L. P. Eisenhart, *Trans. Amer. Math. Soc.*, **26**, 1924, p. 382.

[7] Loc. cit., p. 583.

[8] J. M. Thomas, "On Normal Coördinates for the Geometry of Paths," these PROCEEDINGS, **12**, January, 1926; theorem 1.

[9] H. Weyl, *Raum, Zeit. Materie*, 4th ed., p. 113.

[10] Veblen and T. Y. Thomas, loc. cit., pp. 584–589.

[11] L. P. Eisenhart, *Ann. of Math.*, **24**, p. 373.

[12] L. P. Eisenhart, *Ibid.*, p. 377.

[13] Loc. cit., pp. 599–608.

[14] T. Levi-Civita, *Annali di Math.*, 2d ser., **24**, p. 276.

EINSTEIN'S RECENT THEORY OF GRAVITATION AND ELECTRICITY

BY LUTHER PFAHLER EISENHART

DEPARTMENT OF MATHEMATICS, PRINCETON UNIVERSITY

Communicated December 16, 1925

1. In proposing his recent theory of gravitation and electricity Einstein[1] has derived his equations by expressing that a certain integral is stationary for the variations of a covariant tensor density of the second order and the coefficient of an asymmetrical connection. In this note we show more particularly what kind of a linear connection Einstein has employed and obtain in tensor form the equations which in this theory should replace Maxwell's equations.

2. Let a_{ij} be the components of a covariant tensor of the second order for which the determinant $a = |a_{ij}|$ is different from zero and denote by a^{ij} the cofactor of a_{ij} in a divided by a, then

$$a^{ij}a_{ij} = a^{ji}a_{hi} = \delta^j_k, \tag{2.1}$$

where δ^j_k is 1 or 0 according as $j = k$ or $j \neq k$. We put

$$A^{ij} = \sqrt{ea}\ a^{ij}, \tag{2.2}$$

where e is plus or minus 1 so that the radical is real; as thus defined A^{ji} is a tensor density.

Einstein's equations are (to within a change in notation)

$$L_{jk} \equiv \frac{\partial L^i_{ji}}{\partial x^k} - \frac{\partial L^i_{jk}}{\partial x^i} + L^h_{ji}L^i_{hk} - L^h_{jk}L^i_{hi} = 0, \tag{2.3}$$

and

$$\frac{\partial A^{ij}}{\partial x^k} + A^{ih}L^j_{kh} + A^{hj}L^i_{hk} - A^{ij}L^h_{kh} - \delta^j_k\left(\frac{\partial A^{ih}}{\partial x^h} + A^{hl}L^i_{hl}\right) = 0, \tag{2.4}$$

where L^i_{jk} are the coefficients of the linear connection.

Contracting (2.4) for i and k and j and k, we get

$$\frac{\partial A^{ij}}{\partial x^i} - \frac{\partial A^{ji}}{\partial x^i} = 0, \tag{2.5}$$

$$\frac{\partial A^{ih}}{\partial x^h} + A^{hl}L^i_{hl} = A^{ih}\lambda_h, \tag{2.6}$$

where

$$\lambda_h = \frac{1}{n-1}(L^j_{jh} - L^j_{hj}). \tag{2.7}$$

If in (2.4) we replace A^{ij} by their expressions (2.2) and make use of (2.6), we obtain

$$\frac{\partial a^{ij}}{\partial x^k} + a^{ij}\frac{\partial \log \sqrt{ea}}{\partial x^k} + a^{ih}L^j_{kh} + a^{hj}L^i_{hk} - a^{ij}L^h_{kh} - \delta^j_k a^{ih}\lambda_h = 0. \qquad (2.8)$$

Multiplying by a_{ij}, summing for i and j and making use of the identity

$$\frac{\partial \log a}{\partial x^k} = -a_{ij}\frac{\partial a^{ij}}{\partial x^k},$$

we obtain (since $n \neq 2$)

$$\frac{\partial \log \sqrt{ea}}{\partial x^k} - L^h_{kh} + \lambda_h = 0. \qquad (2.9)$$

Hence equations (2.8) become

$$\frac{\partial a^{ij}}{\partial x^k} + a^{ih}L^j_{kh} + a^{hj}L^i_{hk} - (a^{ij}\lambda_k + \delta^j_k \; a^{ih}\lambda_h) = 0. \qquad (2.10)$$

In consequence of (2.1) equations (2.10) can be written

$$\frac{\partial a_{ij}}{\partial x^k} - a_{ih}L^h_{kj} - a_{hj}L^h_{ik} + (a_{ij}\lambda_k + a_{ik}\lambda_j) = 0. \qquad (2.11)$$

3. We put

$$L^i_{jk} = \Gamma^i_{jk} + \Omega^i_{jk}, \qquad (3.1)$$

where Γ^i_{jk} and Ω^i_{jk} are the symmetric and skew-symmetric parts, respectively, of L^i_{jk}. If we put

$$\Omega_j = \Omega^i_{ij} = -\Omega^i_{ji},$$

we have from (2.7)

$$\lambda_h = \frac{2}{n-1}\Omega_h.$$

If we put

$$B^i_{jkl} = \frac{\partial \Gamma^i_{jl}}{\partial x^k} - \frac{\partial \Gamma^i_{jk}}{\partial x^l} + \Gamma^h_{jl}\Gamma^i_{hk} - \Gamma^h_{jk}\Gamma^i_{hl},$$

and denote by b_{jk} and φ_{jk} the symmetric and skew-symmetric parts of the tensor $B_{jk} = B^i_{jki}$, we have

$$b_{jk} = \frac{1}{2}\left(\frac{\partial \Gamma^i_{ij}}{\partial x^k} + \frac{\partial \Gamma^i_{ik}}{\partial x^j}\right) - \frac{\partial \Gamma^i_{jk}}{\partial x^i} + \Gamma^i_{hj}\Gamma^h_{ik} - \Gamma^i_{ih}\Gamma^h_{jk},$$

$$\beta_{jk} = \frac{1}{2}\left(\frac{\partial \Gamma^i_{ij}}{\partial x^k} - \frac{\partial \Gamma^i_{ik}}{\partial x^j}\right). \qquad (3.2)$$

When the expressions (3.1) are substituted in (2.3), we obtain

$$L_{jk} = b_{jk} + \beta_{jk} + \Omega_{jk}, \tag{3.3}$$

where

$$\Omega_{jk} = \Omega^i_{ji,k} - \Omega^i_{jk,i} + \Omega^h_{ji}\,\Omega^i_{hk} + \Omega^h_{jk}\,\Omega_h.$$

Here, and in what follows, a subscript preceded by a comma indicates covariant differentiation with respect to the Γ's; the expressions for this covariant differentiation are obtained from those of Riemannian geometry by replacing Christoffel symbols of the second kind by Γ's with the same indices.

4. In order to bring his equations into agreement with experimental results, Einstein makes the restriction

$$\lambda_i = \frac{2}{n-1}\Omega_i = 0. \tag{4.1}$$

If we put

$$a_{ij} = g_{ij} + \varphi_{ij}, \qquad a^{ij} = h^{ij} + \psi^{ij}, \tag{4.2}$$

where g_{ij} and h^{ij} are symmetric, and φ_{ij} and ψ^{ij} are skew-symmetric, equations (2.5) become

$$\frac{\partial \sqrt{ae}\,\psi^{ij}}{\partial x^i} = 0. \tag{4.3}$$

Moreover, equations (2.11) are equivalent to

$$g_{ij,k} = \varphi_{ih}\,\Omega^h_{kj} + \varphi_{jh}\,\Omega^h_{ki} \tag{4.4}$$

$$\varphi_{ij,k} = g_{ih}\,\Omega^h_{kj} - g_{hj}\,\Omega^h_{ki} \tag{4.5}$$

From equations (4.5) it follows that the tensor Ω^h_{ij} is given by

$$\Omega^h_{ij} = \tfrac{1}{2}g^{hk}(\varphi_{ij,k} - \varphi_{jk,i} - \varphi_{ki,j}), \tag{4.6}$$

where g^{hk} is defined by

$$g^{hk}\,g_{ik} = \delta^h_i.$$

Hence the conditions (4.1) impose the conditions

$$g^{ik}\,\varphi_{ij,k} = 0. \tag{4.7}$$

From (2.9), (3.1) and (4.1) we have

$$\frac{\partial \log \sqrt{ea}}{\partial x^k} = \Gamma^h_{hk}. \tag{4.8}$$

Consequently the tensor β_{jk} defined by (3.2) is zero, and if we equate to zero the symmetric and skew-symmetric parts of L_{jk}, equations (3.3) are equivalent to

$$b_{jk} + \Omega^h_{ji}\,\Omega^i_{hk} = 0 \tag{4.9}$$

and

$$\Omega^i_{jk,i} = 0. \tag{4.10}$$

5. If in (2.10) we substitute for a^{ij} from (4.2) and make use of (4.1), we get

$$h^{ij}{}_{,k} + \psi^{ir}\Omega^j_{kr} + \psi^{jr}\Omega^i_{kr} = 0 \tag{5.1}$$

and

$$\psi^{ij}{}_{,k} + h^{ir}\Omega^j_{kr} - h^{rj}\Omega^i_{kr} = 0. \tag{5.2}$$

Evidently these equations are equivalent to (4.4) and (4.5).

Contracting (5.2) for i and k, we have

$$\frac{\partial\psi^{ij}}{\partial x^i} + \psi^{rj}\,\Gamma^i_{ir} = 0.$$

Consequently equations (4.3) and (4.8) are equivalent.

Gathering together the preceding results, we see that the field is defined by a symmetric tensor g_{ij}, a skew-symmetric tensor φ_{ij} and a set of functions Γ^h_{ij} symmetric in i and j subject to the conditions (4.4), (4.7), (4.8), (4.9) and (4.10), where Ω^k_{ij} is defined in terms of these functions by (4.6). To these must be added necessarily the conditions of integrability of (4.4).

We can give another form to equations (4.10), if we introduce functions h_{ij} defined by

$$h_{ij}h^{ik} = \delta^k_j.$$

With the aid of these functions, it follows from (5.2) that

$$\Omega^k_{ij} = \tfrac{1}{2}(h_{ir}h_{js}h^{tk}\psi^{rs}{}_{,t} - h_{rj}\psi^{rk}{}_{,i} - \psi^{kr}{}_{,j}h_{ri})$$

and thus equations (4.10) are

$$(h_{ir}h_{js}h^{tk}\psi^{rs}{}_{,t} - h_{rj}\psi^{rk}{}_{,i} - \psi^{kr}{}_{,j}\,h_{ri})_{,k} = 0. \tag{5.3}$$

When we put

$$g_{ij} = -\delta_{ij} + \gamma_{ij},$$

where δ_{ij} is 1 or 0 as $i = j$ or $i \neq j$, and consider γ_{ij} and φ_{ij} as of the first order, the first order terms of equations (4.3) and (4.7) reduce to

$$\frac{\partial\varphi_{ij}}{\partial x^i} = 0$$

and of (5.3) to

$$\sum_k \frac{\partial}{\partial x^k}\left(\frac{\partial\varphi_{ij}}{\partial x^k} - \frac{\partial\varphi_{jk}}{\partial x^i} - \frac{\partial\varphi_{ki}}{\partial x^j}\right) = 0.$$

Thus (4.3), (4.7) and (5.3) are the equations in tensor form which give as first approximation the equations which Einstein proposes as a substitute for Maxwell's equations of empty space. To these must be added also equations (4.4) and (4.9).

When φ_{ij} are identified as the electric and magnetic intensities, they are not the components of the curl of a vector as in the classical theory, unless an additional condition is added.

[1] *Berlin Sitzungsberichte*, **22**, 1925, 414–420.

EVIDENCE FOR THE EXISTENCE OF ACTIVATED MOLECULES IN A CHEMICAL REACTION

BY OLIVER R. WULF

GATES CHEMICAL LABORATORY, CALIFORNIA INSTITUTE OF TECHNOLOGY

Communicated January 14, 1925

It is known that when ozone decomposes rapidly radiation is emitted, part of which at least lies in the visible and ultra-violet. Stuchtey[1] has photographed the spectrum of the radiation using a quartz spectrograph. The radiation is scattered across the visible and extends well out into the ultra-violet. In particular, the well known bands of ozone in the vicinity of 2450 Ångstrom units come in definitely in the spectrum of the luminescence of the decomposition; there is also radiation of still higher frequency extending to approximately 2200Å which seems to have been the limit of observation of Stuchtey's work.[2] It is the purpose of this article to give the probable explanation of the emission of this remarkably short wave-length radiation, which, as will be shown, requires for its excitation energy quantities much larger than can be afforded by the ordinary heat of reaction.

The heat of decomposition of ozone is well known, the best value being that given by Jahn,[3] who found 34,500 calories per mol. The decomposition can under no circumstances be in entirety a simple monomolecular process, for we cannot obtain an even number of oxygen atoms from one of ozone. The interaction of three molecules is impossible as the complete mechanism; for this also does not yield an even number of oxygen atoms. The interaction of four molecules, while leading to possible mechanisms, is so improbable that we shall not consider it here.

We hence conclude that the process leading to the decomposition of ozone involves the interaction of two molecules and thus would make available not more than the energy of decomposition of two molecules. This does not mean that the mechanism which determines the rate of

decomposition is bimolecular, and indeed from our knowledge of the kinetics of this decomposition it may be said with considerable certainty that such is not the case. It does mean, however, that the energy available as a result of a single complete decomposition process will not exceed the energy of decomposition of two molecules of ozone.

The individual process in the case of the decomposition of two molecules of ozone would only liberate $2 \times 34{,}500/6.06 \times 10^{23}$ calories. Expressing this in terms of the frequency of a light quantum of the same energy, or better still in terms of the corresponding wave-length, we have 4120 Å. Hence, if the decomposition results from the interaction of two ozone molecules of normal energy content, and if the energy liberated in decomposition were in certain of the elementary processes liberated as light, the shortest wave-length that could possibly be emitted would be in the blue at 4120 Å.

Actually wave-lengths corresponding to an energy of practically two times this value are observed to be emitted. Now, the spectrum of the emitted light agrees closely with the emission spectrum of ozone molecules. We must accordingly conclude that ozone molecules have been raised from the normal state to states of excitation of much greater energy, the reversion of which again to the normal state being the process of the emission of the observed radiation. Our interest lies in discovering the source of this energy. If we reject the highly improbable possibility of this excitation occurring in steps, then it is a consequence[4] of the experimental facts that at least a part of the *decomposing* ozone molecules be in an activated[5] state, a state of energy-content higher than the normal state, so that the total energy available as the result of the decomposition is the energy represented by *the reversion of two normal ozone molecules to oxygen plus the energy of activation of the reacting molecules*. That is, at least a part of the ozone molecules do not revert to oxygen directly from their normal state; instead these molecules first become activated to states of higher energy content. These molecules may then revert to oxygen, the total energy liberated, a quantity greater than that corresponding to the transformation of two normal ozone molecules into oxygen being available for the excitation of ozone molecules. These excited ozone molecules then constitute the source of the radiation that is observed.

This seems to be a first case where in a homogeneous chemical reaction, there is direct evidence of the formation of activated molecules prior to reaction.

It should be noted that there is still uncertainty as to whether perfectly pure ozone emits radiation when it decomposes.[6] This might be thought to seriously limit the conclusions drawn above, since in a reaction between ozone and impurity there might be sufficient energy liberated to account for the very short radiation that is observed to be emitted. It is possible

that ozone could react with a compound of carbon (for example, acetylene) where the heat of reaction would be sufficient to account for the observed radiation. But it does not appear at all probable that such a possible reaction could occur in a single step. Moreover, the phenomenon has been observed by several investigators under conditions of gas purity which must have varied widely. It appears very probable that the influence of the impurity is a catalytic one, accounting in this way for the bringing out of the luminescence at lower temperatures than those at which it appears in purer ozone. For with increased rate of decomposition the luminescence would be expected to become visible at a lower temperature than that at which it appears in the uncatalyzed reaction. Stuchtey finds no change in the spectrum when it is excited by purely thermal means, or when excited partially by the introduction of an oxidizable impurity, the luminescence spectrum in all cases showing far-reaching agreement with the emission spectrum of ozone. No bands are observed that would have to be ascribed to the impurity. The existence of activated reactants is of course just as necessary a consequence of the catalyzed decomposition as of the normal decomposition, since no more energy is available in the individual process in the former case than in the latter. Therefore the ideas expressed in this paper seem to rest upon grounds of sufficient certainty to allow of the conclusion that in ozone decomposition we have experimental evidence that at least some of the ozone molecules suffer activation before they revert to oxygen.[7]

Summary.—It is shown in this article that the radiation emitted when ozone decomposes into oxygen, considered in relation to the quantum theory, makes it necessary to assume that ozone gas, under conditions favorable for its decomposition, contains ozone molecules in activated states of much higher energy content than that of the normal molecule. The existence of molecules as well as atoms in states of energy content higher than that of the state in which they are normally found, has of course been common experience in the study of the physical processes of the absorption and emission of radiation. But such molecules have hitherto in no case been identified with the activated molecules postulated by Arrhenius to explain the temperature coefficient of chemical reactions. This article is believed to give the first direct evidence which has been offered of the existence of these activated molecules.

The author gratefully acknowledges the suggestions and helpful criticism received from Professors A. A. Noyes and R. C. Tolman and Dr. Linus Pauling.

[1] Stuchtey, *Zs. wiss. Photogr.*, **19**, 161 (1920).

[2] The important question as to whether the oxygen bands in the vicinity of 1900 Å which are active in converting oxygen to ozone also are present in the luminescence spectrum remains as yet unanswered.

[3] Jahn, *Zs. anorg. Chem.*, **60**, 337 (1908). Kailan and John, *Zs. anorg. Chem.*, **68**, 243 (1910).

[4] It must be pointed out that this is a consequence only if we also reject chain reactions in which consecutively normal ozone molecules are decomposed for the activation of some molecule in successive steps, which, after a sufficient number of such steps, may clearly be made to possess enough energy to account for the observed excitation energies. Chain mechanisms of this type, and of such a design as to actually account for the observed radiation appear, from energy considerations and our knowledge of the kinetics of ozone decomposition, to be highly improbable.

[5] By this activation there is understood any passage of the ozone molecule to a form of higher energy. Thus a possible form of activation would be $O_3 \longrightarrow O_2 + O$.

[6] Trautz and Seidel, *Ann. Physik*, **67**, 527 (1922).

[7] In regard to the possibility of reaction between ozone and impurity affording sufficient energy to account for the luminescence observed without any activation of these reactants, it may be pointed out that if it be assumed, in view of the work of Trautz and Seidel referred to above, that it is the reaction between ozone and carbon monoxide which yields the energy that excites the luminescence, the conclusions of this paper remain substantially unchanged. For, while this reaction affords more energy than the reaction of two normal ozone molecules, it does not give a sufficient amount to account for the short wave-length radiation that is observed. Thus from thermochemical data it may readily be computed that the reaction $O_3 + CO = CO_2 + O_2$ is attended by an energy liberation of 102,500 calories. This might account for radiation of wave-lengths as short as 2780Å, but not shorter. As has been said above, Stuchtey finds an important part of the total radiation to be of considerably shorter wave-length than this, namely that lying in the vicinity of 2540Å, and radiation extending even farther in the direction of shorter wave-length than this latter is also observed. Therefore the conclusions of this paper can be applied equally as well to the reaction between ozone and carbon monoxide, if this reaction be thought responsible for the luminescence.

---

*CRYSTALLINE INSULIN**

BY JOHN J. ABEL

FROM THE GATES CHEMICAL LABORATORY, CALIFORNIA INSTITUTE OF TECHNOLOGY AND THE DEPARTMENT OF PHARMACOLOGY, JOHNS HOPKINS UNIVERSITY

Communicated February 6, 1926

In two preceding communications[1,2] it has been told how my collaborators and I were induced to attack the problem of the isolation of insulin and what progress had been made towards its solution to within a few months past. It was there pointed out that a clue had been found, the following out of which promised to lead to the desired goal. This clue eventuated from the observation that when the highly impure and complex, though therapeutically serviceable insulin extracts of commerce are boiled for a short time with $N/10$ sodium carbonate the resulting physiological inactivation of the extracts is always associated with an alteration of a part of their sulphur, an element which our experiments justified us in believing to be an integral constituent of the hormone itself. After

such treatment with a weak alkali a new property appears in the altered insulin in that it now shows an extraordinary sensitivity to very dilute acids which immediately, and contrary to their usual action, liberate hydrogen sulphide from it. It was found that inert fractions prepared from such extracts of the pancreas contain very little of this labile sulphur and a table was constructed which showed that the labile sulphur of a preparation is directly proportional to its hypoglycaemic activity for experimental animals. In other words, the higher the amount of labile sulphur in a given preparation the greater its potency in lowering the percentage of blood sugar.

The methods usually employed in fractioning and purifying biological mixtures have been found to be quite useless in the attempt to separate the hormone from the numerous impurities that are associated with it in even the best of the extracts employed in medical practice. Equally unsuitable to the purpose in mind was the employment of the method of isoelectric precipitation from pH 4.7 to 5.0 no matter how frequently repeated, as likewise the analogous method of electro-dialysis—a method which has not yielded results of any particular value in the hands of those who have tried it out on insulin extracts. We are here dealing with an ampholyte present in very small amount in a solution of other ampholytes whose solubility and precipitation coefficients lie very close to one another. It seemed worthwhile to the writer therefore to employ only weak bases and acids, more especially such as can act as buffers, in the hope that there might result from their use such an alteration of solubilities, that impurities would either fall out of or remain in solution while the hormone in each case took the opposite course. The successful outcome of the experiments proved the essential correctness of the idea.

For a description of our earlier methods of purification the reader is referred to the first paper of this series. The employment of the simple method there described, led uniformly to the preparation of an insulin fraction, Fraction IV, with a rabbit unitage of 40 or more units to the milligram, a unitage that has lately been increased by a more skilful use of these earlier methods. The next steps in the process leading directly to the crystallization of insulin are as follows.

One gram, approximately, of the so-called Fraction IV, is dissolved in a little more than the required volume of $N/6$ acetic acid, enough water is added to bring the volume up to 60 cc. or thereabout, and the contaminating substances (together with some insulin) are then precipitated by the addition of an acidulated solution of brucine containing 6 gms. of brucine in 95 cc. of $N/6$ acetic acid. The resulting clear supernatant fluid which contains nearly pure insulin is separated from the precipitate by centrifugalization. Insulin remaining in the precipitate may be removed by dissolving in $N/6$ acetic acid and precipitating with the brucine solution

as before. How profitable it may be to repeat the process has not been determined at the moment. The clear colorless centrifugalate containing the insulin is then precipitated with $N/6$ pyridine and the precipitate and fluid are immediately centrifuged. The precipitate which settles out is largely crystalline in character, the sides of the tube are found to be coated with glistening highly refractive crystals and the topmost layer of the precipitate consists of similar crystals. It is not a difficult matter to remove this topmost layer of crystals by means of a rubber-mounted

FIGURE 1
Crystals precipitated from acetic acid solution with N/6 pyridine.

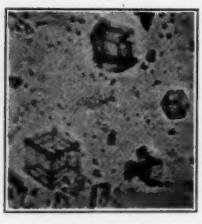


FIGURE 2
Crystals precipitated from Na_2HPO_4 solution with N/6 acetic acid.

Figures 1 and 2. Photo-micrographs, magnification 460 ×, taken by transmitted light.

pipette and to free them from adherent pyridine and acetic acid by frequently washing them with distilled water at room temperature in which medium pure crystalline insulin is quite insoluble. The crystals which are of the type shown in figure 1 may be recrystallized by dissolving them in $N/6$ acetic acid and reprecipitating with $N/6$ pyridine. It has been my practice, however, of late to dissolve this crystalline insulin together with whatever fine granular material there happens to be mixed with it in $M/15$ disodium hydrogen phosphate. To the clear solution $N/6$ acetic acid is added drop by drop with vigorous shaking in order that any insulin thrown out will at once pass into solution again, and this addition of the acid is continued until a slight amount of turbidity persists per-

manently. The flask is then carefully set aside overnight. On the following morning the bottom of the flask as well as its sides are coated with an abundance of colorless, highly refractive crystals. The small amount of amorphous insulin that falls out in very minute granules or floccules—an expression of the above-mentioned turbidity—is readily separated by means of the pipette during the washing with cold water, as the compact crystals obtained by this method sink quickly to the bottom of the tube like so much sea sand. This process of crystallization may be repeated as often as is thought to be necessary. Only a few experiments have as yet been made with the object of crystallizing the compound from hot solvents, and to prepare crystalline derivatives.

Professor Swartz of the Department of Geology of the Johns Hopkins University has kindly examined samples of the insulin crystals shown in figures 1 and 2 and reports them as being very uniform in character suggesting a high degree of purity, as doubly refracting and as belonging to the rhombohedral division of the hexagonal system. The crystals which are allowed to form more slowly out of acidulated disodium hydrogen phosphate solution are very much larger than those precipitated quickly by pyridine out of an acetic acid solution, such crystals being readily perceptible to the naked eye as individual crystals. Irrespective of the manner of their preparation the crystals melt sharply at 233°C. with slight browning occurring at 215°C. The melting point has remained constant on recrystallization.

Chemical Reactions.—A large fraction of the sulphur of the molecule is in the labile state previously described as characteristic of all insulin fractions of high rabbit unitage. Quantitative analyses of both the labile and more firmly bound sulphur are in progress. The compound gives a beautiful biuret reaction passing from a clear pink to a fine purple on further addition of copper sulphate to the alkaline solution. Millon's reaction is unequivocal and positive. The Pauly and the ninhydrin reactions are also positive. This pure insulin is very sensitive to alkali, as shown by the fact that boiling for 15 minutes with $N/10$ sodium carbonate deprives it entirely of its hypoglycaemic activity. It will be seen that the chemical reactions here described are of considerable significance as shedding light on the chemical structure of the compound. It is now being examined in respect to its other physical and chemical properties and efforts will also be made to identify its various cleavage products.

Animal Experiments.—Numerous tests have already been made in order to establish the exact rabbit unitage of the crystalline compound. The rabbit unit employed by us is that fraction of a milligram per kilogram of body weight which will cause convulsions or lower the blood sugar to 0.045 per cent within 5 hours. This corresponds to approximately $1^1/_2$ times the "clinical unit" used at present.

The exact dosage of the crystalline material has not yet been established, but for the present, as is shown in the subjoined table (1), amounts as small as $^1/_{100}$ of a milligram per kilo lowered the blood sugar to about the convulsive level, 0.045 per cent. In another series of experiments $^1/_{125}$ of a milligram per kilo produced convulsions in one rabbit and lowered the

TABLE 1

BLOOD SUGAR DETERMINATIONS

Insulin crystals dissolved in $M/15$ Na_2HPO_4.
1 mgm. equivalent to 10 cc. solution.
Injections were made subcutaneously.

NO.	WT. KILOS	DOSE PER KILO MGM.	ACTUAL DOSE CC.	TIME	BLOOD SUGAR—MGM. PER 100 CC. INITIAL	1½ HRS.	3 HRS.	5 HRS.	UNITS[a]	REMARKS
44	2.07	0.016	0.345	11.23	Lost	50	70	88	60	
39	1.75	0.0125	0.22	11.29	113	32	54	79	80	
42	1.62	0.0125	0.2	11.35	97	50	54	102	80	
15	1.54	0.011	0.17	11.38	113	34	39	47	90	
29	1.51	0.01	0.151	11.43	110	57	92	110	100	
12	1.93	0.01	0.193	11.48	125	48	57	79	100	

TABLE 2

14	1.6	0.008	0.13	11.06	119	25	..	..	125	Convulsion
46	1.95	0.008	0.15	11.11	115	56	77	111	125	

[a] As defined in the text.

blood sugar almost to the convulsive level in a second animal. (Table 2.) It will be clear, therefore, that we are dealing here with a hormone which, in very high dilution, influences carbohydrate metabolism acting as it were like a catalyst in a chemical reaction.

With the isolation of this hormone in crystalline form we are leaving the region of qualitative experiments, and may look forward to quantitative results both in the biological and chemical field which the future promises to yield.

My collaborators in this field of work, Dr. E. M. K. Geiling, Dr. C. A. Rouiller and Dr. David Campbell and I hope to be able in due season to publish more extensive reports on the various researches which naturally arise in the investigation of the properties, chemical and biological, of the pure product.

* Carried out under a grant from the Carnegie Corporation.

[1] Abel, J. J., and Geiling, E. M. K., *Jour. Pharm. and Exper. Therap.*, **25**, 423, 1925.

[2] Abel, J. J., Geiling, E. M. K., Alles, G., and Raymond, A., *Science*, **67**, 169, 1925.